MODEL AND IDENTIFICATION THEORY
FOR
DISCRETE SYSTEMS

By
STANLEY LOUIS SMITH

UNIVERSITY OF FLORIDA

August, 1966

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

.

LIST OF FIGURES

MODEL AND IDENTIFICATION THEORY FOR DISCRETE SYSTEMS

by

Stanley Louis Smith

August 13, 1966

Chairman: Dr. A. P. Sage
Major Department: Electrical Engineering

Discrete-time systems have assumed increasing importance with the
advent of high-speed digital computers. Digital simulation of continu-
ous systems, digital implementation of control strategies, and identi-
fication of discrete systems, requires techniques for accurate model-
ling of system dynamics. Classical methods for discrete representation
of systems continue to be employed while recently introduced methods
offer decided advantages but have not received wide attention. Compre-
hensive studies of both classical and modern methods for discrete
modelling have not appeared to permit evaluation of their relative
merits. An intensive study is undertaken to determine the effective-
ness of the different discretization techniques for the digital simu-
lation of linear and nonlinear continuous systems. A number of digital
experiments are performed to obtain quantitative data for comparison of
the capabilities of the discrete modelling methods.

The improved performance of recently developed methods for digital

simulation is related to certain features of the discretization procedure. These characteristics indicate an approach for the improvement of other discrete representations. This approach requires the identification of parameters within the discrete model to obtain desired response characteristics. Formulation of a procedure for the parameter identification leads to a two-point boundary-value problem. This problem is resolved via a discrete version of the method of quasi-linearization. A procedure for digital computer implementation of this technique is developed and a number of digital experiments performed. Improved discrete models of continuous systems are shown to be obtained with the technique presented. A discrete formulation for the method of differential approximation is presented and the effectiveness of this approach for parameter identification investigated through a series of computer experiments.

CHAPTER 1

INTRODUCTION

Application of digital computers to the analysis and study of dynamic systems requires a discrete formulation of the relationships describing the system behavior; hence every system, whether naturally of discrete-time form, continuous with sampled-data variables, or purely continuous, is represented as a discrete-time system within the digital machine. Methods of numerical analysis are routinely employed in implementing the digital computer solution of differential equations for continuous systems. This approach achieves the system discretization in an indirect manner and may obscure certain properties of interest in the physical system, even while yielding quite accurate solutions to the system differential equations. In a study of system behavior in response to different inputs and nonstationary system parameters, methods for digital computer implementation of the system differential equations have been sought, permitting a discrete model of the physical system to be obtained which will yield accurate representation of the physical system behavior in the physical time domain, i.e. real-time digital simulation.

Discrete modelling techniques developed over the past several years are in general based on standard methods of numerical analysis and related approximate techniques arising in the engineering sciences. While satisfactory digital simulations have long been achieved for such

- 1 -

cases as relatively slow chemical processes, real—time simulation of many dynamic systems was not possible for some time with the digital equipment and computation methods available. With advances in digital hardware design and increased sophistication of computer organization, real—time digital simulation of many complex dynamic processes has been achieved and increasing effort placed on more efficient and more accurate computational techniques. Any desired degree of accuracy within machine capability may be attained by traditional numerical analysis techniques for solution of differential equations [1,2,3] * but the time required to produce the solution is usually prohibitive for real—time simulation even with high—speed machines, and for the majority of systems of interest. Over approximately the last 20 years, many approximate techniques for the discrete representation of continuous systems have been proposed and have been employed for digital simulations with varying degrees of success. Tustin's method, one of the earliest techniques [4] developed within the engineering field, has been perhaps the best accepted approach for digital simulation and is probably the prevalent method in current applications. Most recently, the discrete modelling technique introduced by Fowler [5] has provided a significant advance in the capability of methods for digital simulation, especially for nonlinear systems.

While many papers have appeared in the general literature of engineering and mathematics discussing specific methods for discrete modelling or digital simulation, there have been few attempts to

---

*Numbers in square brackets refer to entries in the references.

compare a number of techniques and to evaluate their relative effectiveness. In a paper on numerical transform calculus, Boxer [6] discusses the fundamental aspects of some of the earlier techniques but gives major emphasis to the Boxer-Thaler approach and results obtained via this method; he makes only general reference to the comparative accuracy of other methods. A more recent study of digital simulation methods has been reported by Fryer and Shultz [7]. In this study are included those methods mentioned by Boxer and other methods later introduced. The authors attempt to compare the effectiveness of the discretization techniques studied by application of each method to the modelling of a specified example with certain other common constraints on the simulation characterizations. The examples considered were those of linear stationary systems, and the simulations were made for a sampling interval time of rather large magnitude in relation to the system dynamics. This study of digital simulation techniques remains the most comprehensive reported in the open literature, to this writer's knowledge. Introduction of the so-called IBM method of digital simulation, Fowler's method, by Hurt [8], reflected research accomplished concurrently with the work of Fryer and Shultz. The depth of experience indicated by the initial reports of Fowler's technique and reflected in subsequent company documents [9] reveals perhaps the major research endeavor in the digital simulation area to this time. The implication of the results obtained with the Fowler method, as reported in the above-referenced papers, is that extensive comparative studies were made of different discrete methods, but no comprehensive supporting data are presented. There have been no published reports of work accomplished with this method from sources other than those cited above.

The study of discrete modelling techniques reported herein is initiated by presenting some of the basic concepts of discrete-time system theory. A state space approach is presented for the description of discrete systems which facilitates formulation of problems for digital computer solution. Following this introduction to discrete-time systems, the more significant of the classical discrete modelling techniques are discussed. The fundamental concepts of each approach to digital simulation are summarized in a procedure for application of each method. Fowler's method of discrete modelling is presented and the procedure for application of the method illustrated in detail. The technique for optimum discrete representation of the integration operation recently reported by Sage and Burt [10] was extended by Sage [11] for modelling of higher-order operators. As yet, neither has significant computational experience with this technique been reported, nor have practicable procedures and computation algorithms been defined. This technique is herein developed and a procedure presented for formulation of the discretization of more general systems.

After discussing the essential features of each of the digital simulation techniques, the methods are employed for the simulation of two example systems. Since all of the techniques presented have been developed for modelling of linear systems, the first example is a second-order linear system for which the step response is sought. Each simulation method is employed to formulate a digital computer algorithm for determination of the desired response. The response of each simulation to a step input is investigated for a number of sampling intervals, and from these data, the relation of the discretization error of

each simulation to change in sampling interval is evaluated, permitting a comparison of the relative effectiveness of the different techniques. The second example considered is that of a second-order nonlinear system. The discretization procedures are here applied to the modelling of the linear portions of the system, and a study made of the simulation sensitivity to sampling interval size for a step input. Additional experiments are performed with the Fowler and optimum discrete approximation methods in which evaluation is made of the simulation performance for sine and ramp function inputs.

The experimental results obtained in the comparative study of digital simulation techniques indicated above, reveal a possible manner in which a discrete model may be improved. The model is tailored for changing sampling intervals and inputs by adjusting selected parameters in the discrete representation. Formulation of this approach leads to a two-point boundary-value problem which is resolved via a discrete form of the generalized Newton-Raphson method, or quasilinerization [12,13,14]. Evaluation of the desired parameters may also be achieved by means of a discrete form of Bellman's differential approximation technique [12,15,16]. This latter approach may also be employed to estimate parameters for initiating the quasilinearization procedure. Procedures for implementation of these techniques are formulated and computational algorithms developed for digital computer solution of the identification problem. Development of this portion of the research comprises the principal effort of the work undertaken.

The approach taken in the discussion of discrete modelling techniques is that of considering a linear transfer function as an operator. The result of a procedure for discretization of a continuous

system transfer function is then a pulse transfer function which will hopefully perform the same operation on an input signal. Such a discrete operation can be implemented on a digital computer by expressing the pulse transfer function in its difference equation form. The resulting equations provide recursive relationships for efficient digital computer representation of a continuous system and hence provide the most probable avenue to real—time digital simulation.

An important class of digital simulation techniques exists in the simulation languages or digital analog simulators. These techniques in general emphasize convenience for the programmer at the expense of computation time. With a simulation for temporary study having no emphasis on real—time operation, convenience and speed in programming is a decided advantage even at the expense of computation time. Despite the significance of simulation languages they will not be discussed here since the present emphasis is on computation techniques which hopefully result in real—time simulation.

DISCRETE APPROXIMATION TECHNIQUES

State—Space Representation of Discrete—Time Systems

Convenient digital implementation of discrete—time models for systems may be achieved through the concepts of state—space representation of discrete—time systems. The principal impetus for this approach to discrete system characterization has been provided by the work of Kalman and Bertram [17], and later, by the work of Zadeh [18]. The more recent publications of Bekey [19] and Freeman [20] are essentially drawn from earlier referenced works, principally those above.

Given the state $\underline{x}(kT)$ of a system at time $t = kT$, the state at time greater than kT for a system input $\underline{u}(kT)$ is given by

$$\underline{x}(\overline{k+1}\ T) = A(kT)\underline{x}(kT) + B(kT)\underline{u}(kT), \tag{2.1}$$

where $\underline{x}$ is an n—vector, $\underline{u}$ is an m—vector, A is an $n \times n$ system operator matrix, and B is an $n \times m$ input weighting matrix. Having the state vector defined for $k = 0$, the system state for $k > 0$ is obtained by repeated application of the recursion formulas of Equation (2.1) to yield

$$\underline{x}(kT) = \prod_{n=0}^{k-1} A(nT)\underline{x}(0) + \sum_{j=0}^{k-1} [\prod_{n=j+1}^{k-1} A(nT)]\ B(jT)\underline{u}(jT). \tag{2.2}$$

The system state transition matrix, or fundamental matrix, is defined by

$$\Phi(k,m) = \prod_{n=m}^{k-1} A(nT), \text{ for } k > m, \qquad (2.3)$$

where

$$\Phi(k,k) = I, \qquad (2.4)$$

I an n x n identity matrix. With this definition of the transition matrix Equation (2.2) may be written as

$$\underline{x}(kT) = \Phi(k,0)\underline{x}(0) + \sum_{j=0}^{k-1} \Phi(k,j+1)B(jT)\underline{u}(jT). \qquad (2.5)$$

For a stationary system with constant matrices A and B, the transition matrix becomes

$$\Phi(k,m) = A^{k-m}, \text{ and } \Phi(k,0) = A^k,$$

so that the state transition equation is written as

$$\underline{x}(k) = \Phi(k)\underline{x}(0) + \sum_{j=0}^{k-1} \Phi(j)B\underline{u}(\overline{k-j-1}\ T). \qquad (2.6)$$

The formulation exemplified by Equation (2.1) in general provides the basis for efficient digital computer algorithms for a discrete representation of system behavior. This characterization permits convenient incorporation of initial conditions, and representation of nonstationary, nonlinear systems through proper definition of the system operator and input weighting matrices. The discrete models for continuous systems derived by the approximation methods to be discussed may be represented by difference equations of the form of Equation (2.1).

Fundamental Concepts of z—transform Theory

Representation of a dynamic system by a digital computer computational algorithm permits the development of a discrete—time system description which approaches the idealized concepts of conventional sampled—data theory. The operation representing an ideal impulse sampler or modulator such as shown in Figure 1 is achieved by the ordinary computation processes within a digital machine. The continuous signal x(t) may be viewed as producing pulse amplitude modulation on the impulse train g(t), or alternately, the impulse train may be viewed as gating a unity gain amplifier to produce output at distinct instants of time. Assuming the impulse train to consist of delta functions uniformly spaced an interval T in time, the impulse train may be represented by

$$g(t) = \sum_{n=-\infty}^{+\infty} \delta(t-nT). \qquad (2.7)$$
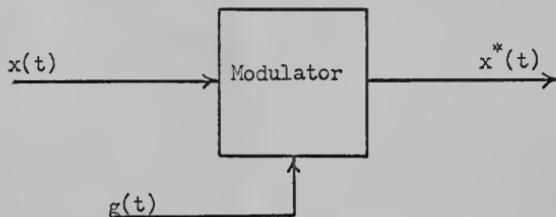
Figure 1. Impulse Modulator—Sampler

With the continuous function x(t) defined for $t \geq 0$, the modulator output $x^*(t)$ is

$$x^*(t) = \sum_{n=0}^{\infty} x(nT) \; \delta(t-nT). \tag{2.8}$$

Taking the Laplace transform of Equation (2.8) results in the form

$$X^*(s) = \sum_{n=0}^{\infty} x(nT)e^{-nTs}. \tag{2.9}$$

Defining the change of variable as introduced by Hurewicz [21], $z = e^{sT}$, leads to

$$X(z) = \sum_{n=0}^{\infty} x(nT)z^{-n}, \tag{2.10}$$

the z-transform of the time function $x(t)$.

Consider the linear, stationary, open-loop, sampled-data system of Figure 2 with synchronized ideal samplers on input and output and impulse response $h(t)$. With the input to $G(s)$ appearing as a sequence

$$\begin{array}{ccccc}
\dfrac{x(t)}{X(s)} & \dfrac{x^*(t)}{X(z)} & \boxed{\quad G(s) \quad} & \dfrac{y(t)}{Y(s)} & \dfrac{y^*(t)}{Y(z)} \\
& T & & & T
\end{array}$$
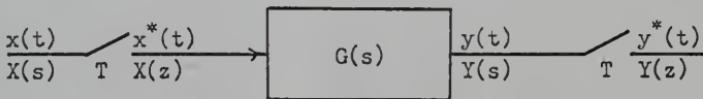
Figure 2. Open-loop Sampled-data System

of impulses, the output $y^*(t) = y(nT)$ at any sampling instant for a relaxed system is given by the convolution summation

$$y(nT) = \sum_{k=0}^{\infty} h(\overline{n-k} \; T)x(kT), \qquad n = 0,1,2,\ldots, \tag{2.11}$$

where the sequence given by the $h(nT)$ is termed the weighting sequence for the sampled-data system and is zero for negative argument. If the z-transform rule of Equation (2.10) is now applied to Equation (2.11), and the index change $j = n-k$ made, the z-transform $Y(z)$ is given by

$$Y(z) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} h(jT)x(kT)z^{-j-k} \qquad (2.12)$$

where $h(jT)$ exists for $j \geq 0$. Recognizing the expression for the z-transform of the weighting sequence and the z-transform of $x(nT)$ imbedded in Equation (2.11), and denoting the weighting sequence transform by $G(z)$, the pulse transfer function, $Y(z)$ is given by

$$Y(z) = G(z)X(z). \qquad (2.13)$$

The broad application of z-transform theory since its introduction by Hurewicz has yielded many comprehensive treatments of the subject [22,23,24] and extensive tabulations of time functions and pulse transfer functions, facilitating application of the z-transform techniques. For analysis problems of great complexity, computer programs are available [9] to aid in performing the z-transform analysis.

Discrete Modelling Methods

Transform Methods

The pulse transfer function resulting from the z-transform operation may be implemented on a digital computer as a difference equation relating the input and output variables of the system under study. Consider a system represented by Figure 2 where the input $x(t)$

and output $y(t)$ are sampled; hence the pulse transfer function may be represented as $G(z) = Y(z)/X(z)$, or

$$G(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-n}}{1 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_q z^{-q}} \quad . \qquad (2.14)$$

Knowing that $z^{-n} F(z) = [f(t-n\,T)]^*$, where $[\ ]^*$ represents the z-transform of the function within the brackets and $F(z) = [f(t)]^*$, leads to a difference equation relating $y(kT)$ and $x(kT)$ when $G(z)$ is replaced by $Y(z)/X(z)$ and the inverse transform obtained for the resulting expression.

This difference equation may now be written as a recursion formula for $y(kT)$,

$$y(kT) = a_0 x(kT) + a_1 x(\overline{k-1}\,T) + \ldots + a_n x(\overline{k-n}\,T) - b_1 y(\overline{k-1}\,T)$$
$$- \ldots - b_q y(\overline{k-q}\,T), \qquad (2.15)$$

which is the desired relationship for digital computer implementation of the discrete approximation resulting from the z-transform method. It is noted that the form of Equation (2.15) is not in the state variable form shown in Equation (2.1). The exact form employed for computer solution depends upon the available computer memory storage.

The z-transform expressions for integration operators, s^{-n}, may be employed in an alternate approach to discretization of a continuous system transfer function. Having expressed a transfer function $G(s)$ as a ratio of polynomials in s^{-1}, substitution is made for the s^{-n} by the corresponding z-transforms. Typical z-transforms for integration operators are given in Table 1. This approach to discretization of a continuous system transfer function is aptly termed integrator

Table 1

Examples of Discrete Forms for Integrating Operators

Methods of Approximation	Operators		
	$1/s$	$1/s^2$	$1/s^3$
z-Transform	$\dfrac{Tz^{-1}}{1 - z^{-1}}$	$\dfrac{T^2 z^{-1}}{(1 - z^{-1})^2}$	$\dfrac{T^3 z^{-1}(1 + z^{-1})}{(1 - z^{-1})^3}$
Tustin	$\dfrac{T(1 + z^{-1})}{2(1 - z^{-1})}$	$\left[\dfrac{T(1 + z^{-1})}{2(1 - z^{-1})}\right]^2$	$\left[\dfrac{T(1 + z^{-1})}{2(1 - z^{-1})}\right]^3$
Madwed–Truxal	$\dfrac{T(1 + z^{-1})}{2(1 - z^{-1})}$	$\dfrac{T^2(1 + 4z^{-1} + z^{-2})}{6(1 - z^{-1})^2}$	$\dfrac{T^3(1 + 11z^{-1} + 11z^{-2} + z^{-3})}{24(1 - z^{-1})^3}$
Boxer–Thaler	$\dfrac{T(1 + z^{-1})}{2(1 - z^{-1})}$	$\dfrac{T^2(1 + 10z^{-1} + z^{-2})}{12(1 - z^{-1})^2}$	$\dfrac{T^3(z^{-1}(1 + z^{-1}))}{2(1 - z^{-1})^3}$

substitution, or internal substitution, and is the approach commonly
taken with several other techniques to be discussed.

Multiplication of the z-transform of s^{-n} by the sampling interval
T is required to obtain the z-transform integration operator. Fryer
and Shultz [7] have reported experimental results for the discreti-
zation of a system utilizing Blum's technique for derivation of digital
filters [25]. The authors observed an unanticipated reduction in the
steady-state gain of the simulation employing this method, but offer
no conclusion regarding the source of the gain loss. It is shown in
Appendix I that the discrete system gain for this approach is reduced
by a multiplicative factor, the sampling interval T, from the
z-transform approximation. This technique for discrete modelling is
mentioned only for its connection with the basic z-transform theory
and to resolve the dilemma posed by Fryer and Shultz; for experimental
results obtained with this approach the reader is referred to the
previously referenced paper [7].

Tustin Method

Originally presented as a general approach to linear system
analysis through the representation of time functions in terms of
sequences of numbers [26], the practical application of the Tustin
method may be reduced to use of Tustin's definition of the differ-
entiating and integrating operators. A linear transfer function $G(s)$
expressed as a ratio of polynomials in s is readily digitized by
substituting for s^n the Tustin operator expressed as

$$s^n = \left[\frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \right]^n$$

where T is the sampling interval. It should be noted that this corresponds to repeated usage of the trapezoidal integration rule.

Madwed–Truxal Method

Madwed extended the Tustin time series approach to system analysis and developed higher order integrating operators of increased accuracy in his comprehensive treatment [27] of this technique. The complex notation developed by Madwed for the polygonal approximation of time functions was clarified by Truxal [28] in his formulation of the numerical convolution operation for system analysis using z–transform notation. The first order integration operator of Madwed is the trapezoidal rule encountered in the Tustin method; however, higher-order Madwed integration operators assume different forms as shown by the examples of Table 1. To digitize a transfer function using this approach, the integrator substitution technique is employed, and the appropriate Madwed integrating operator substituted for s^{-n} to obtain $G(z)$.

Boxer–Thaler Method

The Boxer–Thaler method was presented as a technique for numerical inversion of Laplace transforms [29,6]. The procedure for application of this approach follows that of the previous methods in that substitutions are made for the complex variable s, but the integrating operators employed are the "z–forms" developed by Boxer and Thaler. These special forms were developed in the frequency domain in contrast to the time domain development of Tustin and Madwed. It was noted that polynomial approximation for $s^{-1} = T/\ln z$ could be obtained by expanding ln z in a rapidly convergent series and then expressing the operator s^{-1} as

$$s^{-1} = \frac{T}{\ln z} = \frac{T}{2(u + u^3/3 + u^5/5 + \ldots)} \quad ,$$

where

$$u = \frac{1 - z^{-1}}{1 + z^{-1}} \quad .$$

From this expression there results by synthetic division

$$s^{-1} = \frac{T}{2}(u^{-1} - u/3 - 4u^3/45 - \ldots)$$

which leads to z-forms for s^{-n} when both sides of the above expression are raised to the nth power and the constant term and principal part of the resulting series retained. Table 1 contains z-forms for several orders of integrating operators. A linear system transfer function may be discretized by integrator substitution employing the appropriate z-forms for the s^{-n}.

Anderson-Ball-Voss Method

This technique was presented [30] as an approach to discretization of linear differential equations. If the input to a system can be approximated by a polynomial in time, this approximation is substituted into the system differential equation for the input and the analytical solution determined. For system differential equations of the form

$$A_n\frac{d^n y}{d\,t^n} + A_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \ldots + A_0 y = x(t), \qquad (2.16)$$

the input, $x(t)$, is approximated by a low-order polynomial $h(t)$ permitting the solution, $y(t)$ to be written as

$$y(t) = \sum_{i=1}^{n} c_i e^{a_i t} + H(t)$$

where the a_i are assumed to be distinct. A recursion formula for $y(t)$ can be obtained by writing the solution as

$$y(\overline{m+1}\ T) = \alpha_1 y(mT) + \alpha_2 y(\overline{m-1}\ T) + \ldots + \alpha_j y(\overline{m-j+1}\ T)$$
$$+ \beta_1 x(\overline{m+1}\ T) + \beta_2 x(mT) + \ldots + \beta_{k+1} x(\overline{m-k+1}\ T) \quad (2.17)$$

where j is the order of the differential equation, and

k is the degree of the input approximation polynomial.

Writing the solution $y(t)$ as

$$y(t) = \sum_{i=1}^{n} c_i e^{a_i(t-t_m)} + H(t) , \quad (2.17a)$$

and representing the input approximation by

$$h(t) = h_0 + h_1(t-t_m) + h_2(t-t_m)^2 + \ldots$$

permit the coefficients β_i to be evaluated when $y(t)$ and $h(t)$ are substituted into Equation (2.16) and t made equal to zero. If t is successively made equal to t_m, t_{m-1}, . . ., t_{m-j+1}, a set of j equations are obtained from $y(t)$ for evaluation of the c_i and subsequently the a_i.

Fowler Method

The approach to digital simulation developed by Fowler [31] employs root-locus techniques in conjunction with the z-transforms for the continuous system under consideration. For a nonlinear system such as in Figure 3, the z-transforms of the individual linear transfer functions are first determined.
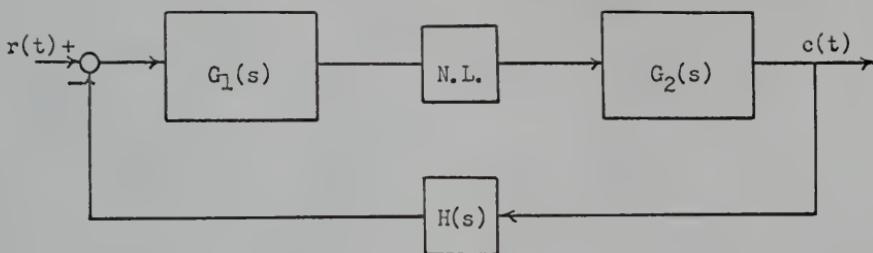
Figure 3. Nonlinear System with a Separable Nonlinearity

The nonlinearity is replaced by a representative gain, frequently
unity, and the poles of the resulting closed-loop pulse transfer
function

$$\frac{G_1(z)G_2(z)}{1 + G_1(z)G_2(z)H(z)} \qquad (2.18)$$

made equal to the poles of the z-transform of the closed-loop
continuous linearized system

$$G(z) = \left[\frac{G_1(s)G_2(s)}{1 + G_1(s)G_2(s)H(s)} \right]^* \qquad (2.19)$$

by adjusting gain parameters in the forward and feedback paths. This
may be accomplished by equating the denominator terms of Equations
(2.18) and (2.19) and solving for the unknown gain parameters. The use
of the notation $F_1(z)$ and $F_2(z)$ in Figure 4 reflects the changed
character of $G_1(z)$ and $G_2(z)$ when the gain parameters are inserted.
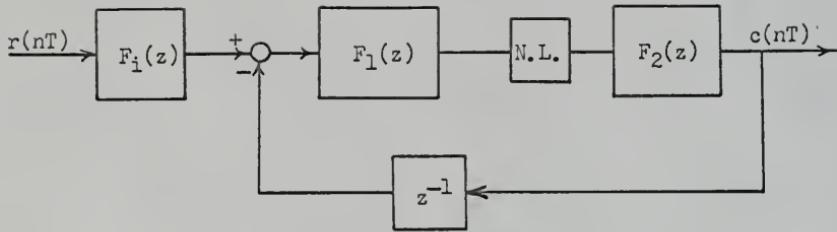
Figure 4. Discrete Model Configuration for the Fowler
Approximation Method

Additional requirements of steady–state gain and system input
approximation are met by determining an input pulse transfer function
$F_i(z)$ of Figure 4 so that the product of this transfer function and the
closed–loop expression (2.18) equals the z–transform of the product of
the desired input approximation or data hold and the linearized closed–
loop transfer function, i.e.

$$\frac{F_i(z)F_1(z)F_2(z)}{1 + F_1(z)F_2(z)} = \left[\frac{H_0(s)G_1(s)G_2(s)}{1 + G_1(s)G_2(s)H(s)}\right]^* \qquad (2.20)$$

where $H_0(s)$ is the data hold transfer function. With the discrete
model completed, it can be implemented by a digital computer program
as shown earlier for pulse transfer functions.

Optimum Digital Simulation

Discrete approximation of a continuous linear system may be
approached by a direct effort to minimize the error between the output
of the discrete system and the sampled output of the continuous system,
as shown by Sage and Burt [10,11]. The signal comparison is made as

illustrated in Figure 5, where r(t) is the input, I(s) is the ideal
continuous operation, H(z) is the unknown pulse transfer function, and
F(z) is a fixed portion of the system.
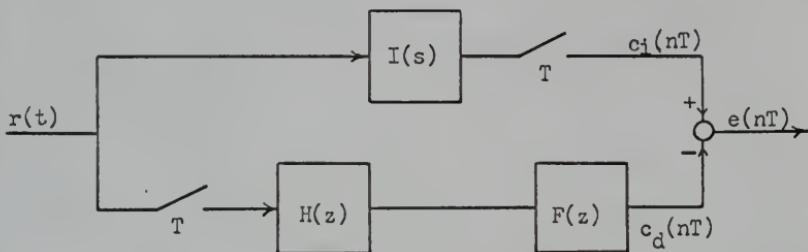
Figure 5. System Configuration for Error Determination

The resulting error sequence is

$$e(nT) = c_i(nT) - c_d(nT)$$

where $c_i(nT)$ is the ideal output after sampling, and $c_d(nT)$ is the
actual discrete system output. Taking the z-transform of $e(nT)$ yields

$$E(z) = [R(s)I(s)]^* - R(z)F(z)H(z). \qquad (2.21)$$

Since the desired approximation is sought to minimize the error above,
the criterion for optimization is chosen as the sum of error squared
which may be expressed as

$$\sum_{n=0}^{\infty} e^2(nT) = \frac{1}{2\pi j} \oint_{\Gamma} E(z)E(z^{-1})z^{-1}dz \quad ,$$

where the contour of integration Γ is the unit circle. Substituting

the expression for E(z) from Equation (2.21) into the integral yields

$$\sum_{n=0}^{\infty} e^2(nT) = \frac{1}{2\pi j} \oint_{\Gamma} [A(z)-R(z)F(z)H(z)] [A(z^{-1}) -$$

$$R(z^{-1})F(z^{-1})H(z^{-1})]z^{-1}dz , \qquad (2.22)$$

where

$$A(z) = [R(s)I(s)]^* .$$

The sum of error squared may be minimized by applying the calculus of variations to the integral of Equation (2.22) yielding the result [32],

$$H_0(z) = \frac{\left\{ \dfrac{R(z^{-1})F(z^{-1})A(z)}{[R(z)R(z^{-1})F(z)F(z^{-1})]_-} \right\}_{P.R.}}{[R(z)R(z^{-1})F(z)F(z^{-1})]_+} , \qquad (2.23)$$

where the symbol P.R. refers to the physically realizable portion of the term within the braces and the + and − subscripts refer to the conventional spectrum factorization operator denoting extraction of the multiplicative term containing poles and zeroes either inside (+) or outside (−) the unit circle.

In the digital simulation of closed—loop systems use of the techniques discussed earlier requires the introduction of a delay in the feedback path to conveniently implement the closed—loop approximation, particularly in the nonlinear case. The need for the delay can be eliminated if transfer functions are approximated in a manner such that computation of the output requires knowledge of only previous values of the output variable. This is achieved in the approach under

discussion by taking the fixed portion of the system as

$$F(z) = e^{-snT} = z^{-n}, n = 1, 2, \ldots \qquad (2.24)$$

If $F(z)$ is taken to be z^{-1}, the pulse transfer function resulting from this approximation technique is determined to give the least sum of error squared when the present value of the dependent variable is not known. Pulse transfer functions with delay are termed closed–loop realizable and those without delay as open–loop realizable. Simulation of a single loop requires only one closed–loop realizable pulse transfer function. Some typical optimum pulse transfer functions are shown in Table 2. The discrete forms for the first order integrator are those developed by Burt [32]; the higher order expressions were derived in the course of this research. It is noted that there is some correspondence to the classical approximations for integrators in Table 1. Sage and Burt [10] have presented a study of the discrete integrator representations.

While system discretization utilizing this approach might be accomplished through integrator substitution, it seems advantageous to obtain the optimum approximation for complete linear transfer functions. Applied to a general system configuration such as illustrated in Figure 3, the method would require obtaining the optimum discrete approximations for the linear transfer functions $G_1(s)$, $G_2(s)$, and $H(s)$. Having determined the discrete model, difference equations can then be written for the system state variables and a computational algorithm developed for a digital computer study of system performance.

Numerical Analysis Methods

The differential equations describing system dynamics may be solved by any of the standard numerical integration methods; however, these methods cannot generally satisfy the requirement for real-time computation and simulation. In the comparison of discrete methods for approximating continuous system response, such integration schemes do provide a convenient means of obtaining results for the continuous system, especially in the nonlinear case. One of the best known approaches to integrating differential equations is the Runge-Kutta method, for which there are many modifications [1,2,3]. A commonly employed formulation is shown here. Given a differential equation

$$\frac{d\underline{x}}{dt} = \underline{f}\,(t,\underline{x})\,,$$

where the sampling interval T is the increment in t, and $\underline{x}$ is an n-vector, a set of coefficients $\underline{a}_i$ are computed where the $\underline{a}_i$ are n-vectors and

$$\underline{a}_1 = T\,\underline{f}(nT,\underline{x}(nT)\,),$$

$$\underline{a}_2 = T\,\underline{f}(nT + \frac{T}{2}\,,\,\underline{x}(nT) + \frac{\underline{a}_1}{2}),$$

$$\underline{a}_3 = T\,\underline{f}(nT + \frac{T}{2}\,,\,\underline{x}(nT) + \frac{\underline{a}_2}{2}),$$

$$\underline{a}_4 = T\,\underline{f}(\overline{n+1}\,T\,,\,\underline{x}(nT) + \underline{a}_3).$$

(2.25)

The new value of $\underline{x}$ is then computed from

$$\underline{x}(\overline{n+1}\ T)\ = \underline{x}(nT) + \frac{1}{6}\ (\underline{\alpha}_1 + 2\underline{\alpha}_2 + 2\underline{\alpha}_3 + \underline{\alpha}_4). \qquad (2.26)$$

This formulation is that of a fourth–order Runge–Kutta method, having a truncation error proportional to T^5. Selection of a sufficiently small increment in the independent variable produces a solution of the desired accuracy; however, the very small increment size sometimes required for a suitable result and the calculation of a complete set of coefficients at every iteration combine to frustrate attempts to utilize such a technique for real–time simulation for most applications.

Experimental Study of the Discrete Modelling Methods

As discussed earlier, few comprehensive, comparative studies on discrete modelling methods or digital simulation techniques have been reported, and none which include more recently introduced methods. Detailed results published for the optimum discrete approximation technique [10,32] consist largely of a study of discrete forms for integrators. Sage [11] has presented a formulation for the optimum approximation approach to digital simulation which included some brief results. For those interested in digital simulation techniques, a comparison of the newer methods with the older, classical approaches to digital simulation may offer some significant insight into progress being made in this area of research. Additional evidence of experience with the modern techniques alone is also of merit. To obtain data for a critical comparison of discrete modelling methods, and to better define the application of the optimum approximation technique, a program of experiments was developed for the digital computer. The

experiments consist of the digital simulation of two example systems
by each of the several discrete modelling methods discussed earlier,
and of determining the simulation sensitivity to changing sampling
intervals and different inputs. Those methods commonly implemented
by integrator substitution are briefly treated in obtaining the
describing difference equations for the example systems, while the
more involved procedures are more completely developed.

Linear System Approximation

Since all the discrete modelling techniques have been developed
primarily for linear transfer function approximation, a fundamental
basis for comparison of the different methods should be their ability
to model a linear system. The example chosen for this purpose is the
second—order linear system shown in Figure 6. The differential
equation describing the system dynamics is

$$\frac{d^2x}{dt} + 6 \frac{dx}{dt} + 25 x = 25 r(t) , \qquad (2.27)$$

which describes a position servomechanism with a damping factor of 0.6
and an undamped natural frequency of 5. radians per second. The
difference equations for simulation of the example system on the
digital computer will be developed for each method. After obtaining
the several representations for the approximating model, the results
for the complete group of experiments will be shown.
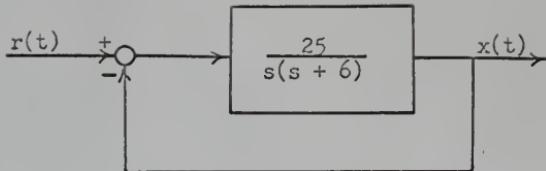
Figure 6. Linear Example System

The system closed—loop transfer function, G(s), given by

$$G(s) = \frac{25}{s^2 + 6s + 25} \quad , \tag{2.28}$$

is placed in the form

$$G(s) = \frac{25s^{-2}}{1 + 6s^{-1} + 25s^{-2}} \tag{2.29}$$

in preparation for obtaining the pulse transfer function by means of integrator substitution.

The Tustin approximation for the example transfer function is obtained through integrator substitution employing the trapezoidal rule integrator form

$$s^{-n} = \left[\frac{T}{2} \frac{1 + z^{-1}}{1 - z^{-1}} \right]^{-n}$$

in Equation (2.29), yielding the pulse transfer function

$$G(z) = \frac{\beta_1 + \beta_2 z^{-1} + \beta_3 z^{-2}}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}} \quad , \tag{2.30}$$

where

$$\alpha_1 = (8 - 50\ T^2)\ \Delta\ , \qquad \alpha_2 = (12T - 25\ T^2 - 4)\ \Delta\ ,$$

$$\beta_1 = 25\ T^2\Delta\ ,\ \beta_2 = 2\ \beta_1\ ,\ \beta_3 = \beta_1\ , \qquad .$$

and $\Delta = (4 + 12\ T + 25\ T^2)^{-1}$.

The desired difference equation can now be obtained in the form of Equation (2.15) as

$$x(\overline{n+1}\ T) = \alpha_1 x(nT) + \alpha_2 x(\overline{n-1}\ T) + \beta_1 r(\overline{n+1}\ T) + \beta_2 r(nT)$$

$$+ \beta_3 r(\overline{n-1}\ T)\ . \qquad (2.31)$$

The Madwed–Truxal form for the pulse transfer function approximating $G(s)$ of Equation (2.28) is obtained by integrator substitution in Equation (2.29) making use of the integrator forms

$$s^{-1} = \frac{T}{2} \frac{1 + z^{-1}}{1 - z^{-1}}$$

and

$$s^{-2} = \frac{T^2}{6} \frac{1 + 4z^{-1} + z^{-2}}{(1 - z^{-1})^2}\ .$$

The resulting pulse transfer function is given by

$$G(z) = \frac{\beta_1 + \beta_2 z^{-1} + \beta_3 z^{-2}}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}}\ , \qquad (2.32)$$

where

$$\alpha_1 = (12 - 100 \, T^2) \, \Delta \, , \quad \alpha_2 = (18 \, T - 25 \, T^2 - 6) \, \Delta \, ,$$

$$\beta_1 = 25 \, T^2 \Delta \, , \quad \beta_2 = 4 \, \beta_1 \, , \quad \beta_3 = \beta_1 \, ,$$

and

$$\Delta = (6 + 18 \, T + 25 \, T^2)^{-1}.$$

The related difference equation then assumes the form of Equation (2.31) with the α_i and β_i as defined here.

Integrator substitution utilizing the z-forms of Boxer and Thaler in Equation (2.29) requires substitution of

$$s^{-1} = \frac{T}{2} \frac{1 + z^{-1}}{1 - z^{-1}}$$

and

$$s^{-2} = \frac{T^2}{12} \frac{1 + 10z^{-1} + z^{-2}}{(1 - z^{-1})^2} \quad ,$$

yielding the pulse transfer function

$$G(z) = \frac{\beta_1 + \beta_2 z^{-1} + \beta_3 z^{-2}}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}}$$

for which

$$\alpha_1 = (24 - 250 \, T^2) \, \Delta \, , \quad \alpha_2 = (36 \, T - 25 \, T^2 - 12) \, \Delta \, ,$$

$$\beta_1 = 25\ T^2\Delta\ , \quad \beta_2 = 10\ \beta_1\ , \quad \beta_3 = \beta_1\ ,$$

and

$$\Delta = (12 + 36\ T + 25\ T^2)^{-1}. \tag{2.33}$$

The Boxer—Thaler approximation may then be completed by implementing
the difference equation of Equation (2.31), with the α_i and β_i of
Equation (2.33).

Since the procedure for application of the Anderson—Ball—Voss
method is somewhat lengthy, and has been stated in complete form
earlier in this chapter, only the principal steps and results are
offered here. The system input time function, $r(t)$, in Equation
(2.27) is replaced by an approximating polynomial of second—order

$$r(t) \equiv K_1 + K_2(t-nT) + K_3(t-nT)^2 \tag{2.34}$$

where $\quad K_1 = r(nT), \qquad K_2 = \dfrac{r(n+1\ T) - r(n-1\ T)}{2T}$

and $\quad K_3 = \dfrac{r(n+1\ T) - 2r(nT) + r(n-1\ T)}{2T^2}$.

The solution to Equation (2.27) with $r(t)$ as defined by Equation
(2.34) may now be written in the form of Equation (2.17a), and the
necessary coefficients evaluated by substituting the solution into
Equation (2.27), with the $r(t)$ of Equation (2.34), and then succes—
sively imposing the conditions $t = nT$, $t = \overline{n-1}\ T$ on the solution.
The difference equation for the solution to Equation (2.27) is now
obtained in the form of Equation (2.17) by letting $t \to \overline{n+1}\ T$, and by

employing the coefficients evaluated above, so that

$$x(\overline{n+1}\ T) = \alpha_1 x(nT) + \alpha_2 x(\overline{n-1}\ T) + \beta_1 r(\overline{n+1}\ T) + \beta_2 r(nT)$$

$$+ \beta_3 r(\overline{n-1}\ T)\ , \qquad\qquad (2.35)$$

where $\qquad \alpha_1 = 2e^{-3T} \cos 4T\ , \qquad \alpha_2 = -e^{-6T}$

$$\beta_1 = (1-\alpha_1-\alpha_2) \frac{11-75\ T}{625\ T^2} + (1+\alpha_2) \frac{25\ T - 12}{50\ T} + \frac{1}{2}(1-\alpha_2)\ ,$$

$$\beta_2 = (1-\alpha_1-\alpha_2) \frac{625\ T^2 - 22}{625\ T^2} + \frac{12}{25\ T}(1+\alpha_2) - (1-\alpha_2)\ ,$$

and

$$\beta_3 = (1-\alpha_1-\alpha_2) \frac{75\ T^2+ 11}{625\ T^2} - (1+\alpha_2) \frac{25\ T + 12}{50\ T} + \frac{1}{2}(1-\alpha_2).$$

For a completely linear system, the application of Fowler's method reduces to the derivation of the z—transform for the linear system transfer function with some desired input data hold or input approximation. If the input data hold for the present example is taken as a zero—order data hold, $H_0(s)$, with an adjusting lead of one—half sample period to compensate for the lag of the basic hold operation, the pulse transfer function sought to model $G(s)$ of Equation (2.28) is

$$[H_0(s)\ e^{.5sT}G(s)]^*\ .$$

Carrying out the indicated z—transformation results in the pulse transfer function

$$G(z) = \frac{\beta_1 + z^{-1}\beta_2 + z^{-2}\beta_3}{1 - z^{-1}\alpha_1 - z^{-2}\alpha_2} \quad ,$$

for which the related difference equation is

$$x(\overline{n+1}\ T) = \alpha_1 x(nT) + \alpha_2 x(\overline{n-1}\ T) + \beta_1 r(\overline{n+1}\ T) + \beta_2 r(nT)$$

$$+ \beta_3 r(\overline{n-1}\ T) , \qquad (2.36)$$

where $\alpha_1 = 2e^{-3T}\cos 4T ,\qquad \alpha_2 = -e^{-6T} ,$

$$\beta_1 = 1 - e^{-1.5T}(\cos 2T + .75 \sin 2T) ,$$

$$\beta_2 = e^{-1.5T}(\cos 2T + .75 \sin 2T) +$$

$$e^{-4.5T}(\cos 2T - .75 \sin 2T) - \alpha_1 ,$$

and $\beta_3 = -\alpha_2 - e^{-4.5T}(\cos 2T - .75 \sin 2T) .$

Since the test input function to be employed in the computer implementation of the equations developed here is a unit step function, the optimum discrete approximation is derived for this input. Desiring an open-loop realizable approximation for the closed-loop transfer function of Equation (2.28), the fixed portion of the discrete model $F(z)$ of Equation (2.24) is made unity, $F(z) = F(z^{-1}) = 1$, and the term $[R(z)\ R(z^{-1})\ F(z)\ F(z^{-1})]$ of Equation (2.23) is $[(1 - z)(1 - z^{-1})]^{-1}$. Factoring this term as required by the procedure to utilize Equation (2.23) results in

$$[R(z) R(z^{-1}) F(z) F(z^{-1})]_+ = \frac{1}{1 - z^{-1}}$$

and

$$[R(z) R(z^{-1}) F(z) F(z^{-1})]_- = \frac{1}{1 - z} \quad .$$

The term $A(z)$ of Equation (2.23) is given by

$$A(z) = \left[\frac{25}{s(s^2 + 6s + 25)} \right]^* \quad .$$

The expression for the optimum pulse transfer function then becomes

$$H_0(z) = \frac{[A(z)] \text{ P.R.}}{R(z)} \quad ,$$

and, since $A(z)$ is physically realizable, $H_0(z)$ results immediately as

$$H_0(z) = \frac{z^{-1}\beta_1 + z^{-2}\beta_2}{1 - z^{-1}\alpha_1 - z^{-2}\alpha_2}$$

where

$$\alpha_1 = 2 e^{-3T}\cos 4T , \qquad \alpha_2 = -e^{-6T} ,$$

$$\beta_1 = 1 - e^{-3T}(\cos 2T + .75 \sin 2T) ,$$

and

$$\beta_2 = e^{-6T} - e^{-3T}(\cos 2T - .75 \sin 2T).$$

The requisite difference equation for the discrete model output is

given by

$$x(\overline{n+1}\ T) = \alpha_1 x(nT) + \alpha_2 x(\overline{n-1}\ T) + \beta_1 r(nT) + \beta_2 r(\overline{n-1}\ T) . \tag{2.37}$$

Solution of the difference equations derived above for the several approximation methods was carried out by programming them for the IBM 1401-709 data processing system. The organization of the programs developed is shown by the flow charts of Appendix IV. A comparison of the discrete models developed for the example system was made by determining the response of the model to a unit step function input at a number of sampling intervals. The approximate model response was compared to the solution of the system differential equation obtained by a fourth-order Runge-Kutta integration method for a small sample interval, .001 second. The standard of comparison for the different methods was chosen as a normalized sum of error squared, NSES, between the approximate model response and that obtained via the Runge-Kutta integration scheme. The NSES criterion is given by

$$\text{NSES} = \frac{1}{N-1} \sum_{k=0}^{N-1} \left\| \underline{x}_i(kT) - \underline{x}_d(kT) \right\|^2 R(kT) , \tag{2.38}$$

where $\underline{x}_i(kT)$, an n-vector, is the continuous system state vector at $t = kT$, the n-vector $\underline{x}_d(kT)$ is the discrete model state vector at the same sample instant, $R(kT)$ is an $n \times n$ positive semi-definite weighting matrix, and N is the number of sample points over the observation interval. The observation interval for the computer experiments was taken as five seconds. An analytical determination of the sum of error squared in Equation (2.38) is possible for linear systems through the

integral relationship of Equation (2.22) for the case where the
weighting matrix, R, is constant and where $N = \infty$. A bilinear trans-
formation for a mapping from the z plane into a w plane,

$$z = \frac{1 + w}{1 - w} ,$$

maps the contour of integration Γ , the unit circle, into an inte-
gration about the left half of the w plane. The requisite integration
in the w plane can be accomplished by employing the integral tables
found in Appendix E of Newton, Gould, and Kaiser [33]. The NSES
criterion is stated in general form here for convenient application to
systems other than that of the immediate example. For the present
examples the weighting matrix is the n x n identity matrix.

Results of the digital computer experiments for sampling intervals
ranging from .01 to 0.3 seconds are shown in Figure 7. The classical
approximate methods and the Fowler method yield quite the same error in
the simulation result for sampling intervals within what would normally
be a practical magnitude. In order to adequately display the system
dynamic behaviour for inputs with high frequency content, the sampling
interval would probably be chosen not greater than 0.2 of the system
rise time, and perhaps as small as 0.1 of the rise time. In this range
of sampling interval size, the majority of the discrete models yield
essentially the same result for the step input. For sampling periods
greater than 0.1 second, the methods become increasingly distinctive in
response, though still not greatly different, with exception of the
Boxer-Thaler approximation, which demonstrates increasing sensitivity
to change in the sampling interval and was found to be unstable for a
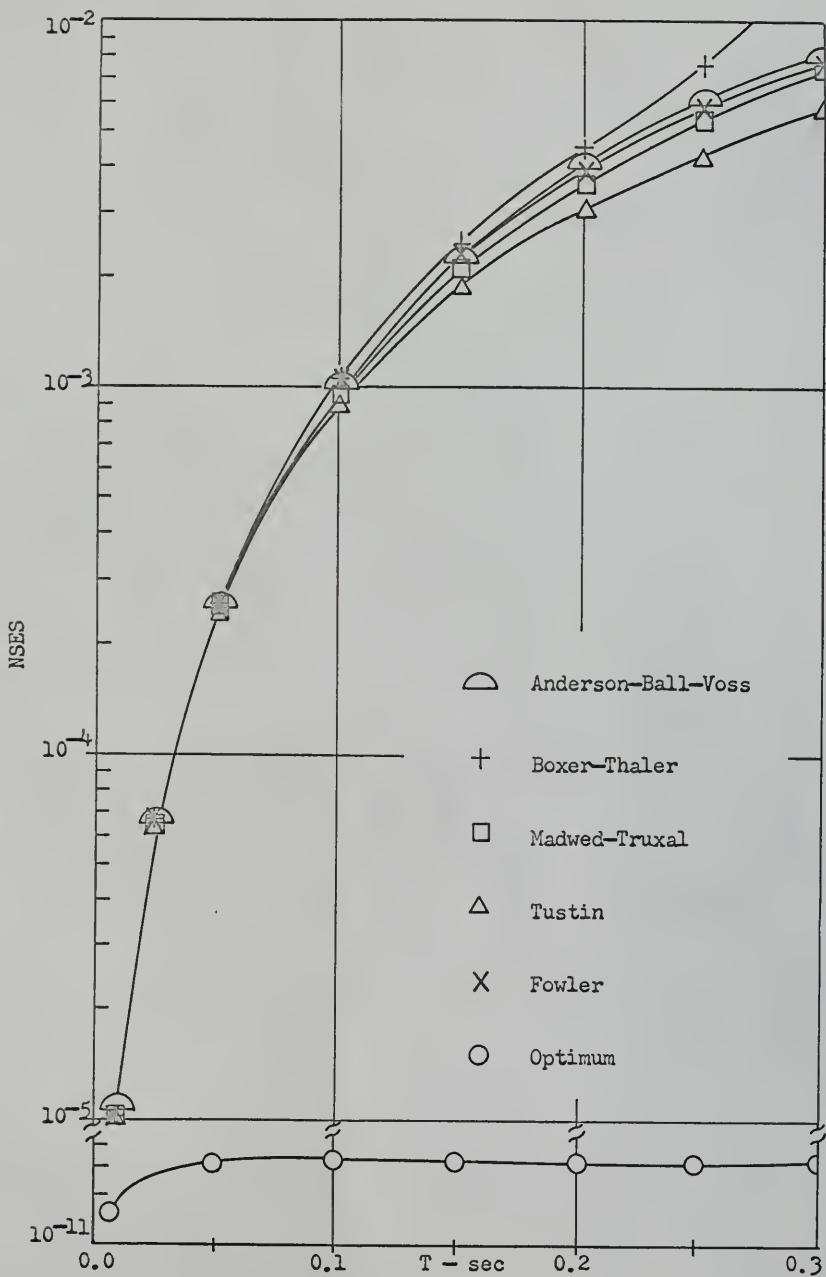
Figure 7. Error Criterion for the Output of the Linear System

sampling period of 0.5 second. Quite distinguishable from the tightly grouped methods of higher error is the optimum approximation error curve of Figure 7. It is noted that the vertical scale is broken, for convenience in plotting, and that the error for the optimum discrete model is lower than for the other methods by a factor of 10^{-6}. Also of importance is the insensitivity of the model to change in sampling interval.

Figure 8 illustrates the step response of the continuous system and the discrete approximations. For the sampling interval of 0.1 second employed for the case shown, the response of the models determined by the Tustin, Madwed-Truxal, Boxer-Thaler, Anderson-Ball-Voss, and Fowler methods are nearly indistinguishable on the drawing scale except for the initial values. The Tustin model response is chosen as representative of the group for improved clarity in the presentation. The initial values for the different methods which are the principal differences in the responses at the stated sampling interval are shown in Table 3. The lack of delay in the forward path of the approximate models for these methods contributes a major portion of the simulation error which distinguishes these techniques from the optimum method. As the sampling interval increases, the initial value increases very rapidly, with an accompanying detrimental effect on the simulation accuracy.

Further comparison of the linear discrete approximations was made for selected methods by computing the response of each method to a ramp function input, the ramp function having a five to one slope. Of the discrete models developed previously, the experiment was conducted for the Tustin, Fowler, and optimum representations. The results are
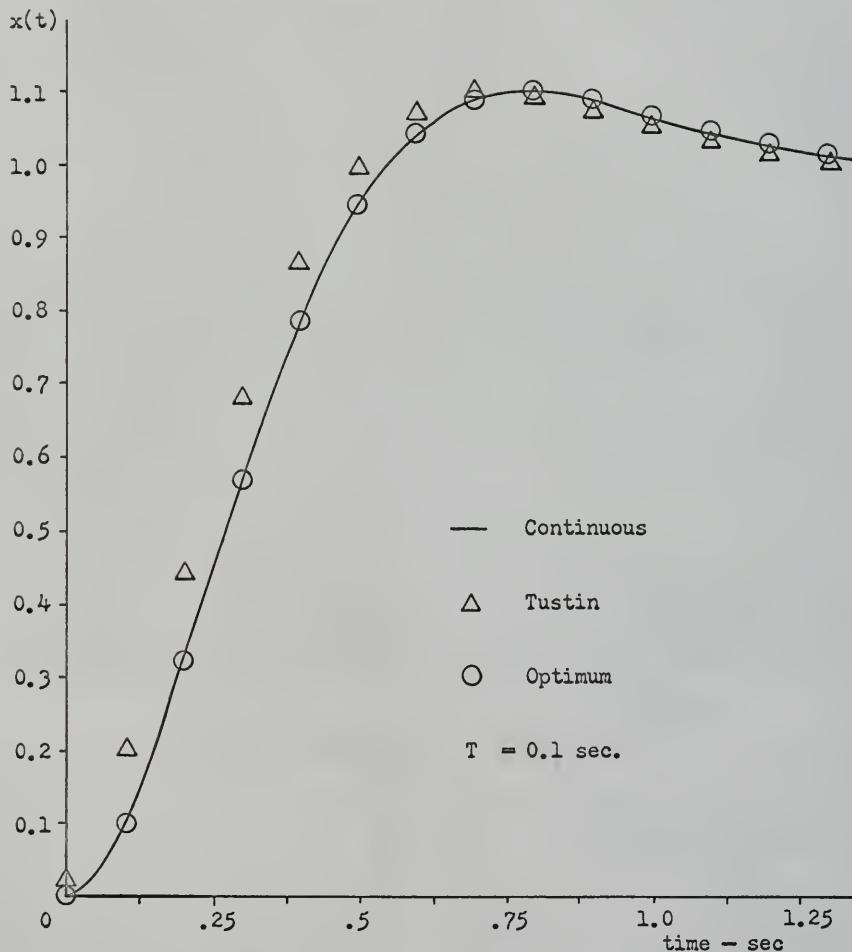
Figure 8. Linear System Unit Step Response

Table 3

Computed Initial Conditions for the
Linear System Simulations

Method	x(0)
Tustin	.04587
Madwed—Truxal	.03106
Boxer—Thaler	.01577
Anderson—Ball—Voss	.02026
Fowler	.02820
Optimum	.00000
Continuous	.00000

displayed in Figure 9 where the data for the response of the Fowler
model are the same as for the Tustin model within the resolution of the
plot. It is apparent that the model which was optimum for a unit step
input is not so for the ramp input. The unit delay in the optimum
discrete model for a step input which permitted a realistic approxi-
mation of physical system delay now appears as a damaging parameter. A
second linear optimum approximation was developed to be optimum for a
ramp input, with the response also shown in Figure 9. This discrete
approximation is developed in Appendix III as a detailed example of the
optimum approximation procedure. For some applications all the models
tested here might be acceptable in performance, but it is clear that
the simulation optimized for the low order input exhibits less accurate
performance with higher order inputs. The values of the NSES criteria
for the discrete approximations employed in the ramp response experiment
are given in Table 4.

Table 4

Normalized Error Square Criterion
for the Ramp Response Experiment

Method	NSES
Tustin	4.2×10^{-5}
Fowler	8.2×10^{-6}
Optimum for step	2.6×10^{-2}
Optimum for ramp	5.8×10^{-6}
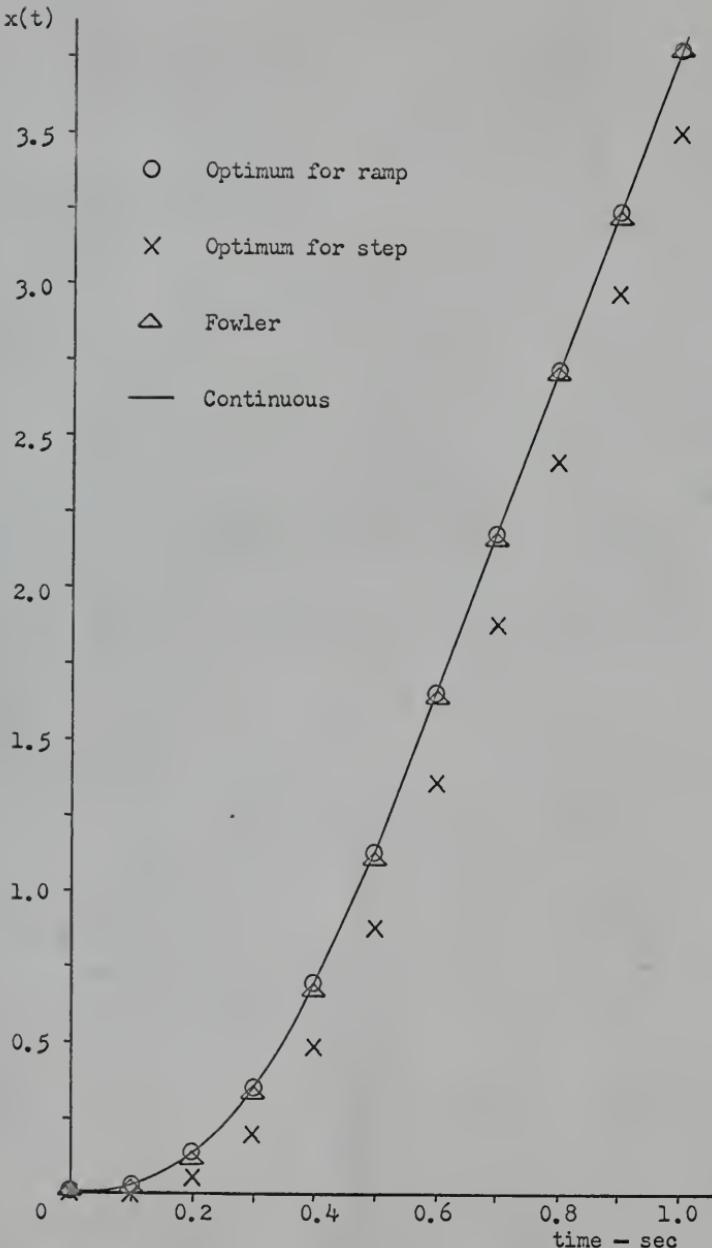
Figure 9. Linear System Ramp Response

Nonlinear System Approximation

Digital simulation and discrete modelling of physical systems can
seldom be accomplished for the majority of modern analysis work through
purely linear approximations. While some segments of complex systems
may certainly be adequately represented by means of a linear model, it
is unlikely that a complete analysis could in general be satisfactorily
completed on this basis. Digital simulation of aircraft dynamics has
been an active area of inquiry and has stimulated the efforts for more
accurate discrete modelling of nonlinear systems [34,35,36]. The air-
craft systems contain nonlinear elements in such forms as amplifier
response and actuator characteristics which must be included in a model
intended for adequate system representation. Consideration is here
given to the discrete modelling of a nonlinear system in order that
additional insight may be gained into the usefulness of methods for
discrete approximation of physical systems. The example system is
shown in Figure 10.
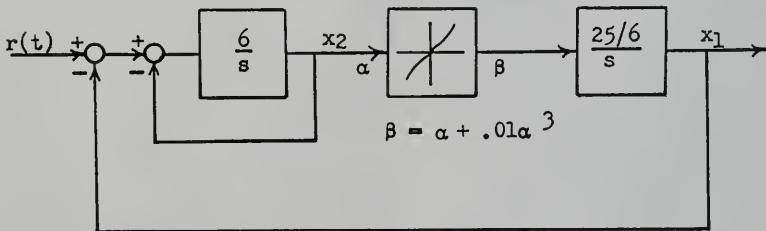
Figure 10. System for Nonlinear Example

The state equations for the example system may be given utilizing the state variable identification of Figure 10 as follows

$$x_1 = \frac{25}{6} [x_2 + .01 \, x_2^3] \qquad (2.39)$$

and

$$x_2 = 6 \, [r - (x_1 + x_2)] \quad ,$$

thus permitting immediate application of the Runge–Kutta integration method for first-order differential equations.

Discretization of the system is best effected by first replacing the minor loop by its closed-loop form. The integrator substitution methods are then applied to the transfer functions

$$G_1(s) = \frac{6}{s+6} \quad , \quad \text{and} \quad G_2(s) = \frac{25/6}{s}. \qquad (2.40)$$

Since the integrator substitution here requires only the first-order integrator form, the Tustin, Madwed–Truxal, and Boxer–Thaler techniques yield the same discrete model. This common formulation of the discrete representation will be termed the Tustin approximation. Substitution of the trapezoidal rule integration operator into $G_1(s)$ and $G_2(s)$ of Equations (2.40) yields pulse transfer functions

$$G_1(z) = \frac{3 \, T(1 + z^{-1})}{(1 + 3T) - (1 - 3T) \, z^{-1}}$$

and

$$G_2(z) = \frac{25}{12} \, T \, \frac{1 + z^{-1}}{1 - z^{-1}} \quad . \qquad (2.41)$$

The discrete model for the example system then appears as shown in Figure 11.
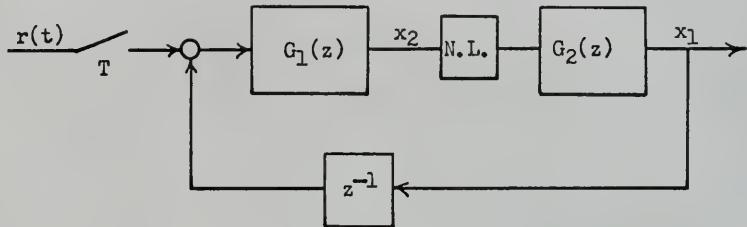
Figure 11. Tustin Model for the Nonlinear System

The difference equations for the state variables of the system may now be written as follows:

$$x_1(\overline{n+1}\ T) \;=\; x_1(nT) + \frac{25}{12}\ T\ [f(\overline{n+1}\ T) + f(nT)]$$

$$x_2(\overline{n+1}\ T) \;=\; \frac{1 - 3T}{1 + 3T}\ x_2(nT) + \frac{3T}{1 + 3T}\ [e(\overline{n+1}\ T) + e(nT)]$$

$$f(nT) \;=\; x_2(nT) + .01\ x_2^3(nT)$$

$$e(nT) \;=\; r(nT) - x_1(\overline{n-1}\ T) \;. \tag{2.42}$$

Development of the difference equations for the Anderson–Ball–Voss approximation for the nonlinear system requires obtaining of the discrete forms of the solutions to Equations (2.39). This is accomplished by carrying out the requisite procedure for each differential

equation as was done for the linear system previously treated in this chapter. The input function approximation employed for the procedure is the linear time function

$$r(t) \equiv r(nT) + \frac{r(\overline{n+1}\ T) - r(\overline{n-1}\ T)}{2T}\ (t-nT)\ . \tag{2.43}$$

The difference equations obtained from this procedure describe a discrete model of the form of Figure 11, and may be stated as

$$x_1(\overline{n+1}\ T) = x_1(nT) + \frac{25}{24}\ T\ [f(\overline{n+1}\ T) + 4\ f(nT)$$

$$+ f(\overline{n-1}\ T)]\ ,$$

$$x_2(\overline{n+1}\ T) = \alpha x_2(nT) + \beta_1 e(\overline{n+1}\ T) + \beta_2 e(nT)$$

$$+ \beta_3 e(\overline{n-1}\ T)\ ,$$

where $f(nT)$ and $e(nT)$ are defined as in Equation (2.42) above, and

$$\alpha = e^{-6T}\ , \qquad \beta_1 = \frac{1}{2} - \frac{1}{12T}\ (1 - \alpha)\ ,$$

$$\beta_2 = 1 - \alpha\ , \qquad \beta_3 = -\beta_1\ . \tag{2.44}$$

The procedure for the Fowler method is shown more explicitly by the development of the discrete model for the nonlinear system than by the previous linear system approximation with the method. The z-transforms for the transfer functions of Equations (2.40) are obtained as

$$G_1(z) = \frac{6}{1 - e^{-6T}z^{-1}}, \quad \text{and} \quad G_2(z) = \frac{25}{6(1 - z^{-1})}.$$

Multiplication of $G_2(z)$ by the sampling interval T is effectively replacing the integrator of the continuous system by a backward difference rectangular integration rule. The numerator of $G_1(z)$ is replaced by a parameter K which is to be adjusted so that the eigenvalues of the closed-loop discrete system correspond to the eigenvalues of the closed-loop linearized continuous system. With these modifications $G_1(z)$ and $G_2(z)$ become the $F_1(z)$ and $F_2(z)$ of Figure 4 and appear as

$$F_1(z) = \frac{K}{1 - e^{-6T}z^{-1}}, \quad \text{and} \quad F_2(z) = \frac{25T}{6(1 - z^{-1})}. \tag{2.45}$$

Letting the nonlinearity of the system be replaced by a simple unity gain, the z-transform for the resulting closed-loop linear system is

$$G(z) = \frac{z^{-1}25\, e^{-3T}\sin 4T}{1 - z^{-1}2e^{-3T}\cos 4T + z^{-2}e^{-6T}}. \tag{2.46}$$

Similarly linearizing the discrete system, the closed-loop pulse transfer function is

$$F(z) = \frac{KT(25/6)}{1 - z^{-1}(1 + e^{-6T} - \frac{25}{6}KT) + z^{-2}e^{-6T}} \tag{2.47}$$

where $H(z)$ in Figure 4 is z^{-1}. The requirement that the closed-loop eigenvalues of the two discrete systems be equal is met if the denominators of Equations (2.46) and (2.47) are the same. This condition is

realized if

$$\frac{25}{6} KT - 1 - e^{-6T} = -2 e^{-3T}\cos 4T ,$$

which yields

$$K = \frac{6}{25T} (1+e^{-6T} - 2e^{-3T}\cos 4T). \qquad (2.48)$$

The final determination for the Fowler model is the input transfer function $F_i(z)$ of Figure 4. This element is sought so that when multiplied times the $F(z)$ of Equation (2.47) the result will yield

$$G(z) = \left[H_0 (s) e^{\cdot 5st}G(s)\right]^* , \qquad (2.49)$$

where $H_0(s)$ is a zero-order data hold transfer function. Development of the z-transform indicated by Equation (2.49) gives

$$G(z) = \frac{\beta_1 + z^{-1}\beta_2 + z^{-2}\beta_3}{1 - z^{-1}\alpha_1 - z^{-2}\alpha_2} ,$$

where

$$\alpha_1 = 2e^{-3T}\cos 4T , \qquad \alpha_2 = -e^{-6T}$$

$$\beta_1 = 1 - e^{-1.5T}\Delta_1 , \qquad \beta_2 = e^{-1.5T}\Delta_1 + e^{-4.5T}\Delta_2 - \alpha_1$$

$$\beta_3 = -\alpha_2 - e^{-4.5T}\Delta_2 ,$$

$$\Delta_1 = \cos 2T + .75 \sin 2T ,$$

$$\Delta_2 = \cos 2T - .75 \sin 2T. \qquad (2.50)$$

The input transfer function requirement is met if

$$\frac{25}{6} KT \, F_i(z) = \beta_1 + z^{-1}\beta_2 + z^{-2}\beta_3 \, , \qquad (2.51)$$

where the β_i are defined in Equations (2.50). Incorporating the evaluation of K from Equation (2.48) gives

$$F_i(z) = \mu_1 + z^{-1}\mu_2 + z^{-2}\mu_3$$

where

$$\mu_1 = \beta_1/D \, , \quad \mu_2 = \beta_2/D \, , \quad \mu_3 = \beta_3/D \, ,$$

and

$$D = 1 - \alpha_1 - \alpha_2 \, . \qquad (2.52)$$

Information is now complete to permit expression of the difference equations for the Fowler approximation to be written, so that

$$x_1(\overline{n+1} \, T) = x_1(nT) + \frac{25}{6} T \, [x_2(\overline{n+1} \, T) + .01 \, x_2^{3}(\overline{n+1} \, T)] \, .$$

$$x_2(\overline{n+1} \, T) = e^{-6T}x_2(nT) + Ke(\overline{n+1} \, T) \, ,$$

and

$$e(\overline{n+1} \, T) = \mu_1 r(\overline{n+1} \, T) + \mu_2 r(nT) + \mu_3 r(\overline{n-1} \, T) - x_1(nT) \, , \qquad (2.53)$$

with the μ_i and K as defined above.

The difference equations for the optimum discrete approximation to the nonlinear system are readily developed through utilization of the discrete forms of Table 2. The integrator of the continuous system

is modelled by the closed—loop realizable integration rule developed for a ramp input, i.e.

$$G_2(z) = \frac{25\ T}{12}\ \frac{z^{-1}(4 - z\ z^{-1} + z^{-2})}{1 - z^{-1}}\ . \tag{2.54}$$

The minor loop is approximated by the open—loop realizable form derived for a ramp input which appears as

$$G_1(z) = \frac{(6T - 1 + e^{-6T}) + z^{-1}(1 - e^{-6T} - 6T\ e^{-6T})}{6T\ (1 - e^{-6T}z^{-1})}. \tag{2.55}$$

The requisite difference equations may be written immediately as

$$x_1(\overline{n+1}\ T) = x_1(nT) + \frac{25}{12}\ T\ [4f(nT) - 3f(\overline{n-1}\ T) + f(\overline{n-2}\ T)]\ ,$$

$$x_2(\overline{n+1}\ T) = \alpha x_2(nT) + \beta_1\ e(\overline{n+1}\ T) + \beta_2 e(nT)\ ,$$

where

$$e(nT) = r(nT) - x_1(nT)\ ,$$

$$f(nT) = x_2(nT) + .01\ x_2{}^3(nT)\ ,$$

$$\alpha = e^{-6T}\ ,$$

$$\beta_1 = (6T - 1 + \alpha)\ /\ 6T\ ,$$

and
$$\beta_2 = (1 - \alpha - 6T\alpha)\ /\ 6T\ . \tag{2.56}$$

Having developed the difference equations needed to implement the discrete models as digital computer programs, a series of computer experiments was conducted similar to those undertaken for the linear example. The sampling interval sensitivity of the different discrete models was determined by computing the step response for each model at a number of different sampling intervals. The results of this sequence of experiments is shown in Figures 12 and 13, where the NSES criterion defined in Equation (2.38) has been computed for sampling intervals in the range 0.01 second to 0.3 seconds for each state variable. It is clear that the performance of the classical methods has deteriorated from that achieved for the linear system approximation. With the test input for this experiment being a 10 unit step function, the system is driven so that the effect of the nonlinearity is evident although not dominating. The Anderson—Ball—Voss approximation becomes unstable very rapidly with increase in the sampling interval being unstable for a sampling interval of 0.09 second. This is attributable in part to the low order input approximation employed in deriving the discrete model for this method. The Tustin approximation becomes unstable at a sampling interval of 0.225 second. Response of the different simulations to a 10 unit step function input is shown in Figures 14 and 15 for a sampling interval of 0.1 second which seems a maximum reasonable value in relation to the example system dynamics. The approaching insta-bility of the Tustin approximation is evident from Figure 14. Results from these experiments also indicate that the improved approximation of the output state by the optimum discrete model is due largely to the more accurate integration operator employed in that model.
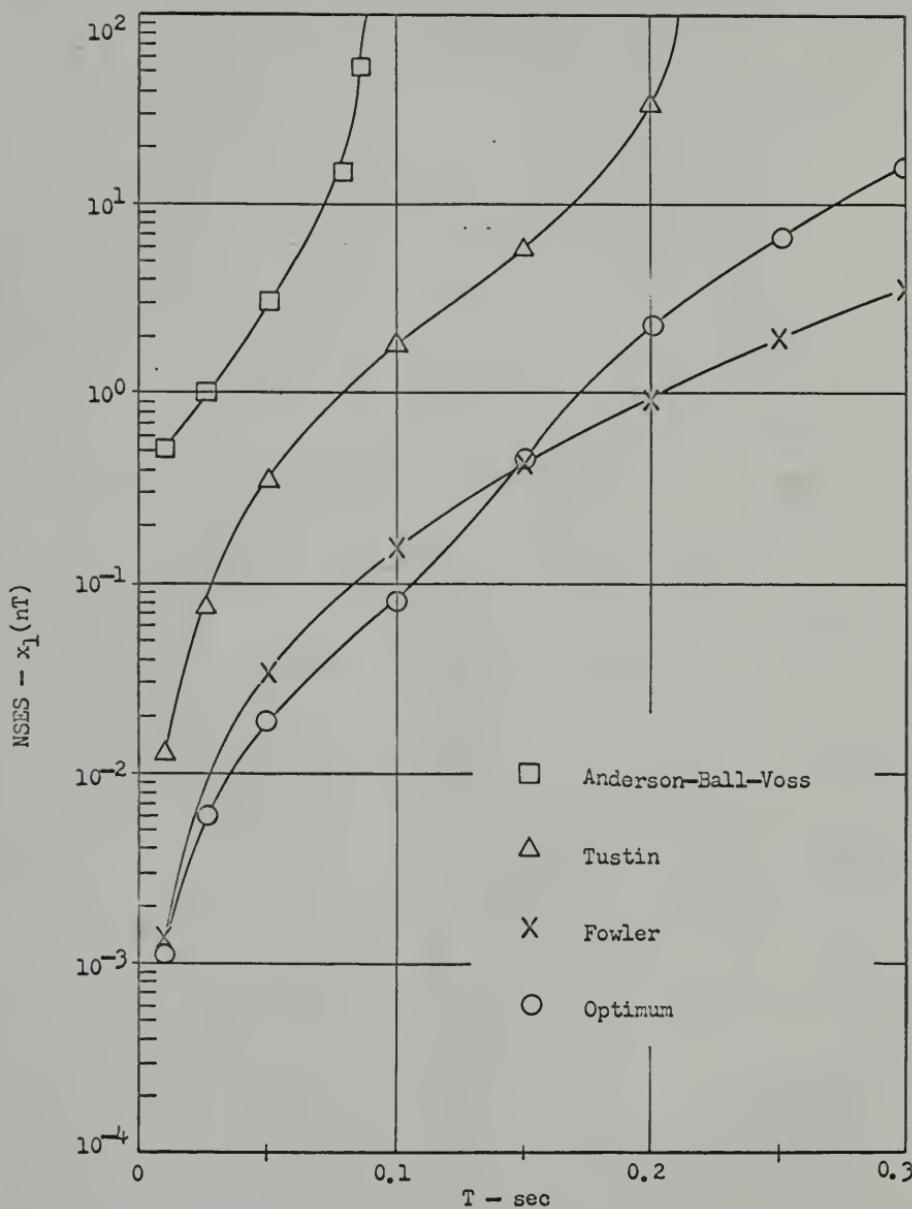
Figure 12. Error Criterion for x_1 with 10 Unit Step Input
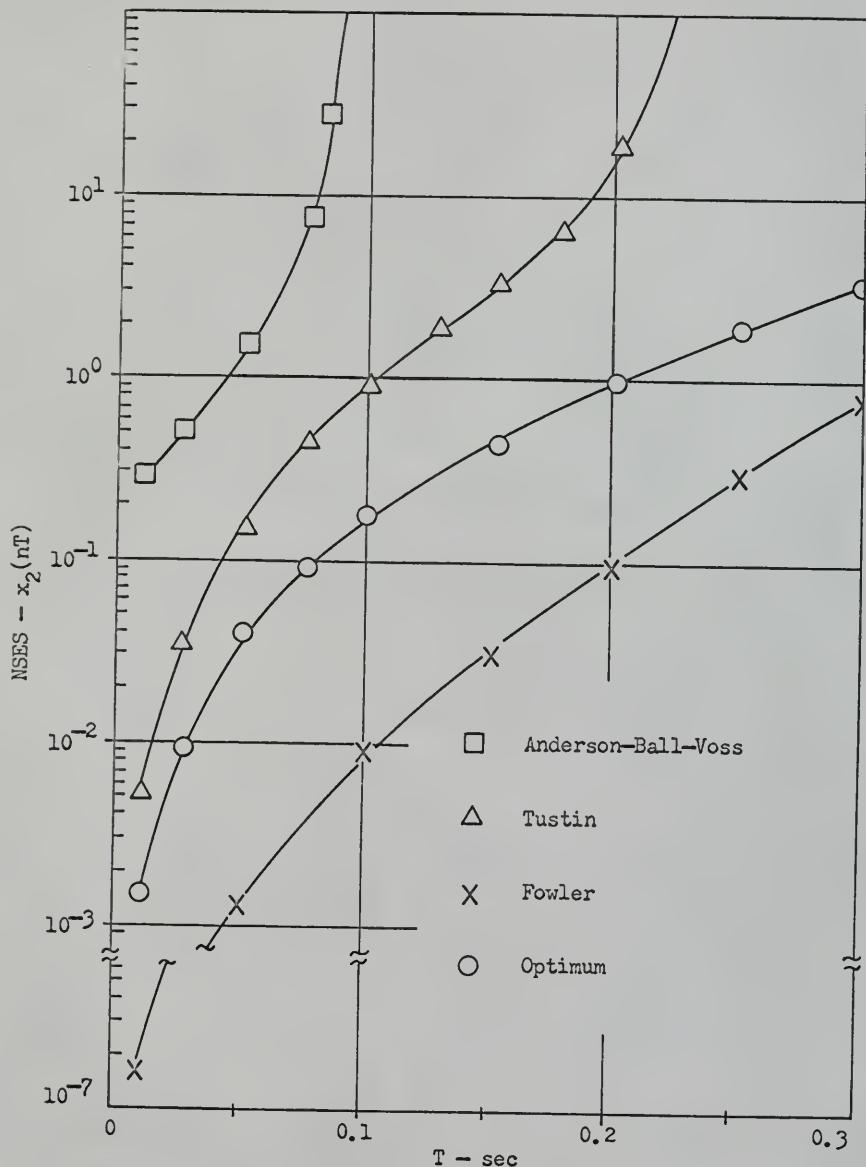
Figure 13. Error Criterion for x_2 with 10 Unit Step Input
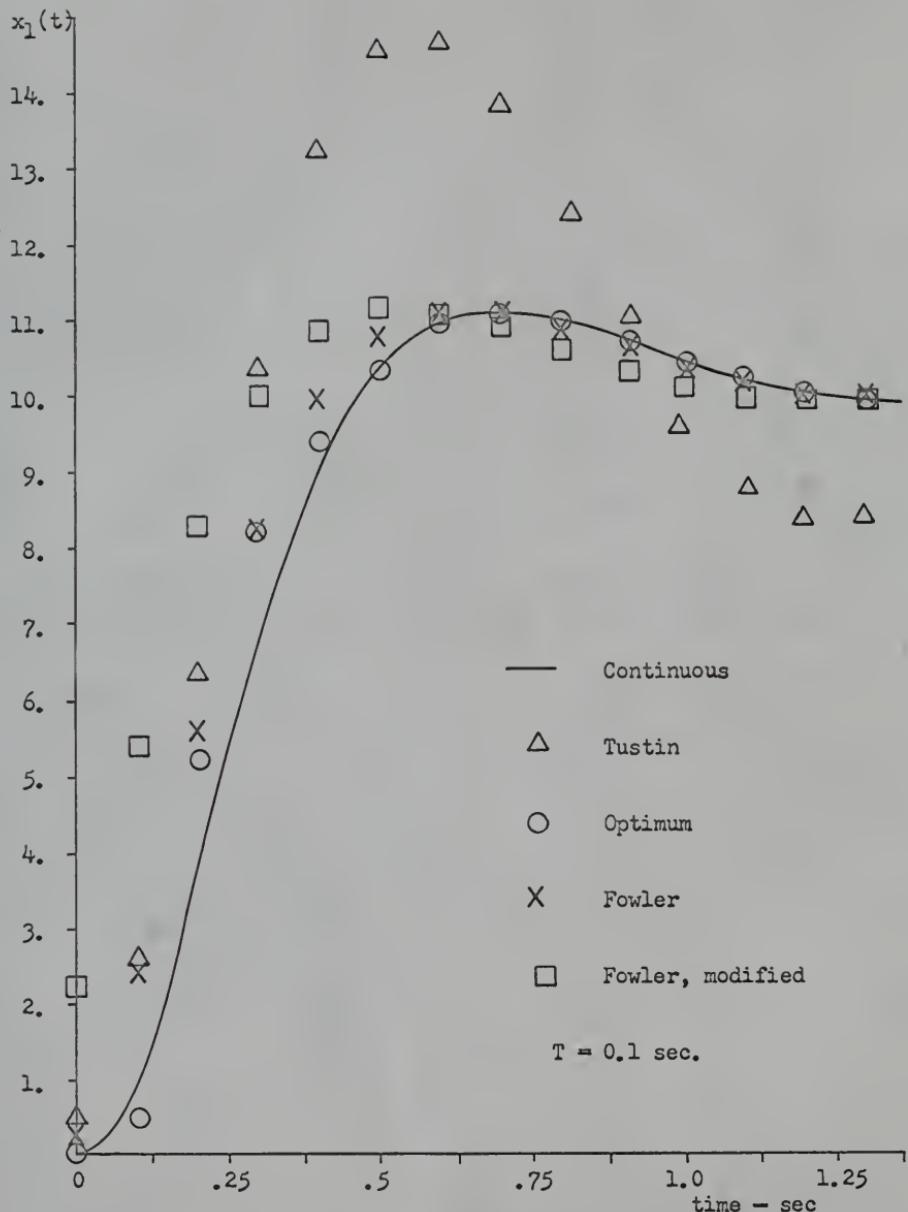
Figure 14. Nonlinear System Response to 10 Unit Step Input
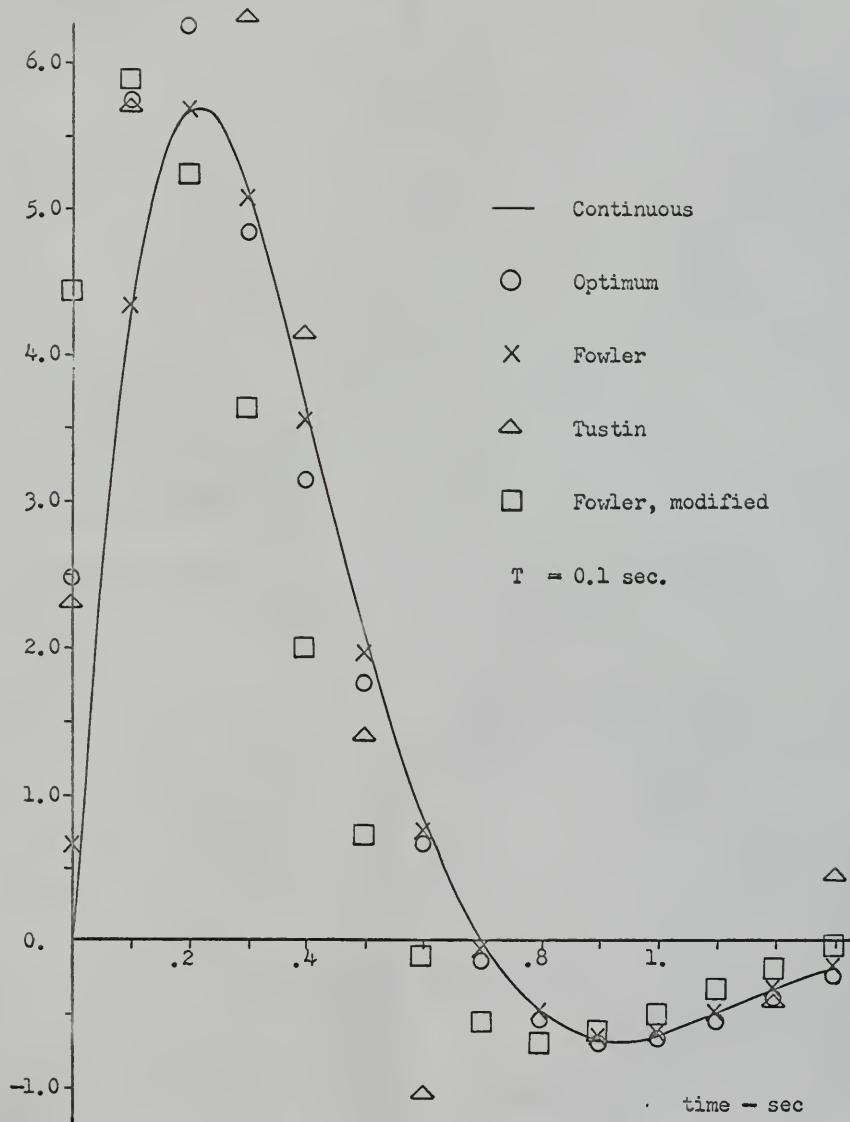
Figure 15.   Nonlinear System $x_2(t)$ Response for 10 Unit Step Input

Perhaps the most significant result of the initial experiment is the evidence in Figure 12 of the lower sensitivity of the Fowler approximation to changes in sampling interval. Although the optimum discrete approximation is indeed optimum for an adequate range of sample interval size, the sensitivity of this model to change in sampling interval is greater than for the Fowler model and is a damaging characteristic. A principal distinguishing feature of the Fowler model is the input transfer function. Since input data holds were not generally treated as a part of all the methods considered here, but only for the Fowler method for which the procedure explicitly in- cludes this feature, it is of interest to observe the performance of a modified Fowler model in which the input transfer function is not present. Such a model is readily derived from the original Fowler approximation, Equations (2.53), by replacing the error term in the equations by

$$e(\overline{n+1} \; T) \;=\; r(\overline{n+1} \; T) - x_1(nT) \; . \qquad (2.57)$$

This modified Fowler approximation was programmed for the series of computer experiments conducted previously. Figures 16 and 17 display the values of the sum of error square for the response of this model to the 10 unit step function input for a study of sampling period sensitivity. The results shown emphasize the significance of a proper input approximation in any discrete model, a fact long recognized. These results do however weaken the implication above that the input transfer function might be the determining factor in the sampling interval sensitivity of the model, for while the overall accuracy of the simulation is decreased by deletion of this part of the model, the
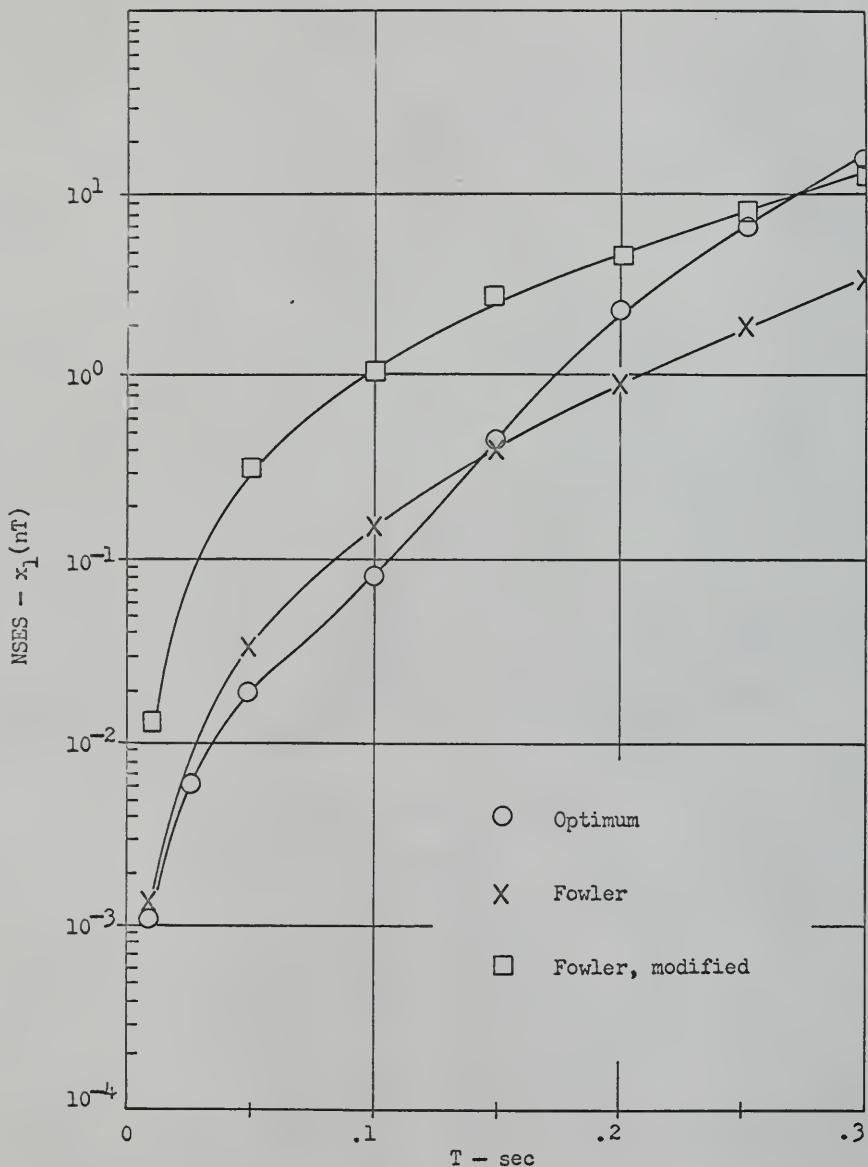
Figure 16. Error Criterion for x_1 of Nonlinear System for Step Input with Modified Fowler Result
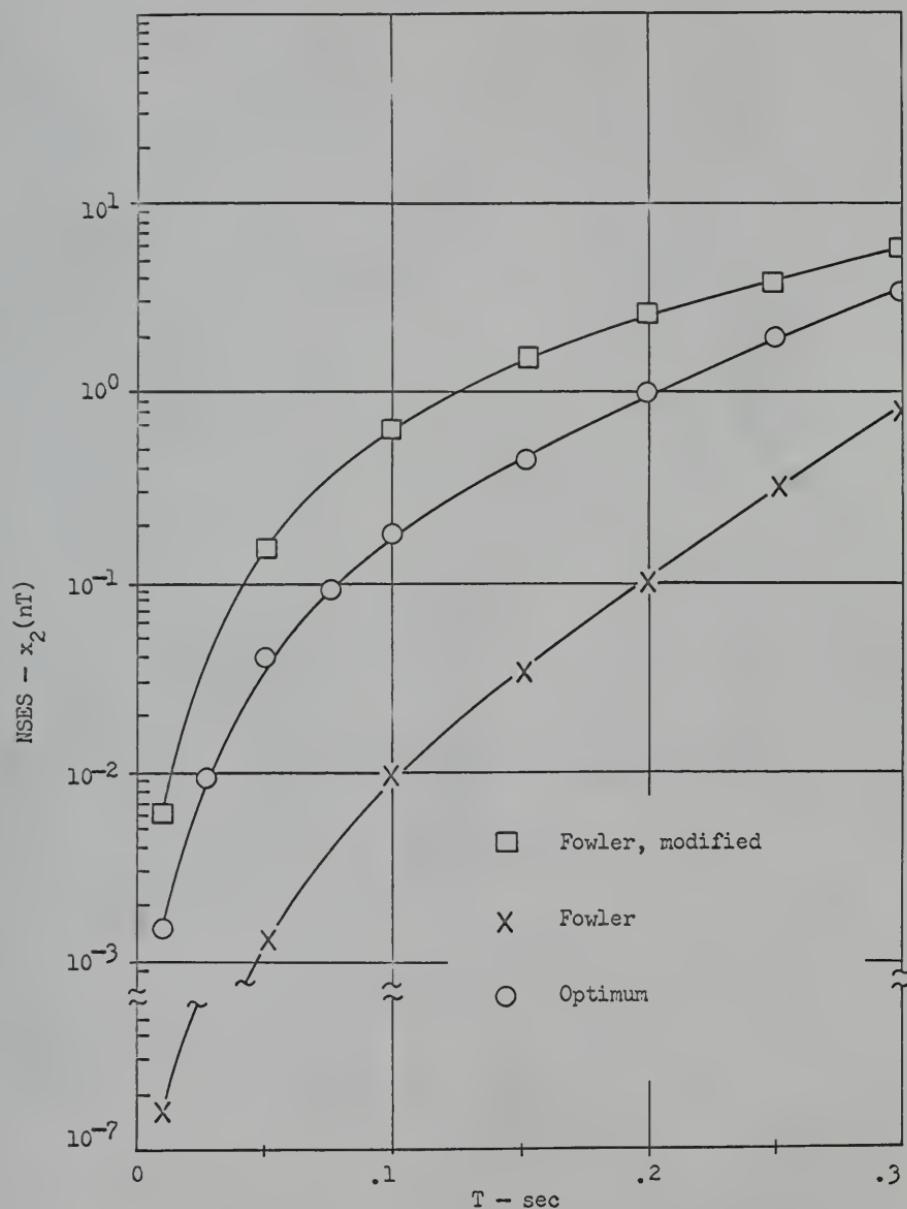
Figure 17.   Error Criterion for x_2 of Nonlinear System for Step Input with Modified Fowler Result

sensitivity to sampling interval change is approximately the same, as evidenced in Figure 16. The assertion by Fowler of the importance of matching the discrete model eigenvalues to those of the continuous system appears to be the critical factor in the modelling procedure for this characteristic. The computational algorithm written for the complete Fowler model is made to achieve the matching of the eigen-values at each new sampling interval for constant improvement of the model dynamics. This feature of the Fowler method suggests that similar adjustment of parameters might be of value in the other methods. Such a scheme for refinement of the optimum discrete approximation will be considered in the next chapter.

Discrete model response to sine wave inputs is a common test of simulation performance and is a characteristic of importance in practical application of discrete systems. Response of the Fowler simulation, complete and modified, and response of the optimum simu-lation to such an input have been determined in a computer experiment. A typical sine wave response of these models shown in Figure 18 reveals the excellent ability of the complete Fowler model and optimum model to simulate the system. The effect of removing the input transfer function from the Fowler model appears principally as a time lead in the dis-crete representation, a characteristic observable also in Figure 14 for the step response. That this time lead is an undesirable character-istic is also evident from the plots of the NSES criterion in Figures 19 and 20. Sensitivity of the different simulations revealed by the NSES data is not so disparate as for the step input response. The apparent reduction in the NSES criterion of the complete Fowler simu-lation for an increase in sample interval at small interval size is a

result of the definition of the criterion. The total simulation error in this region actually doubles, but the normalizing factor, the number of sample intervals, is now 0.2 its original value for a sampling interval of 0.05 second; consequently the normalized sum of error squared appears to be reduced. Although the NSES criterion for x_1 reflects badly upon the optimum discrete approximation for all values of sample interval with the sine wave input, the actual simulation result is not at all objectionable. The optimum discrete model response shown in Figure 18 for a sample interval of 0.2 second is quite acceptable for many applications, and for smaller sampling intervals such as would normally be used in simulating this system, the optimum discrete simulation output response cannot be distinguished from the Fowler model response on a scale of resolution such as that of Figure 18. The results for the state variable x_2 in Figures 20 and 21 give additional support to the capability of the optimum approximation.

For a ramp function input to the example system, the discrete simulations by the Fowler and optimum approximations yield excellent results as shown in Figure 22. The modified Fowler simulation possesses the characteristic lead in the response and accompanying high simulation error. Evaluation of the NSES criterion for the optimum simulation state vector yields 7.2×10^{-5}, for the Fowler simulation 1.5×10^{-3}, and for the modified Fowler simulation .22.

Discrete modelling via direct z-transform replacement of linear system transfer functions has not been experimentally studied. Every discrete model represents in effect a z-transform approximation, and the Fowler method is considered to be z-transform modelling at its best. The modified Fowler model studied does in fact yield results
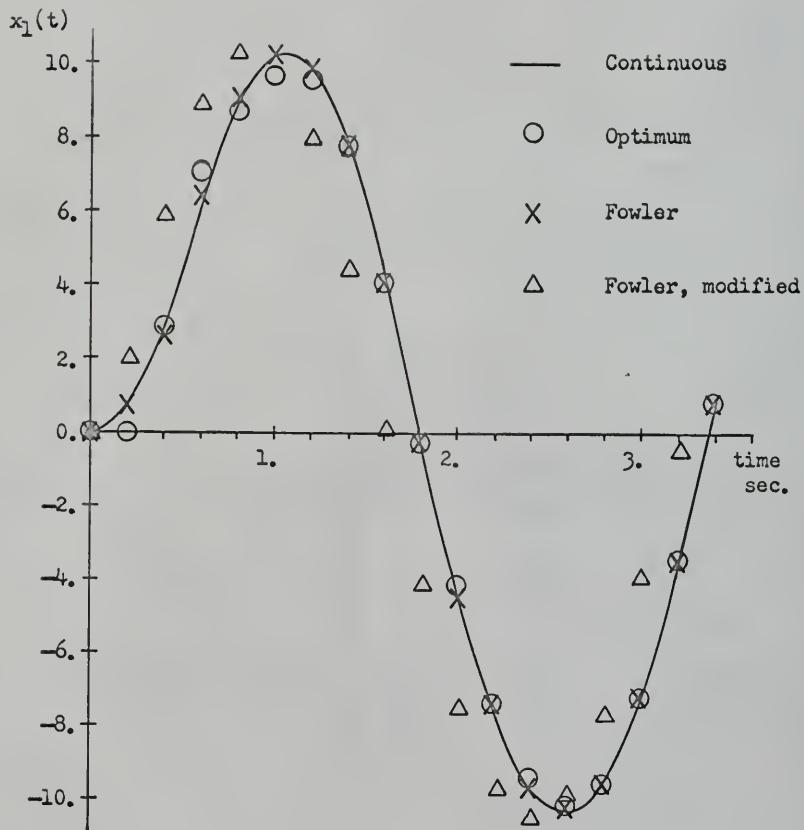
Figure 18. Sine Wave Response of Nonlinear System for T = 0.2 Sec.
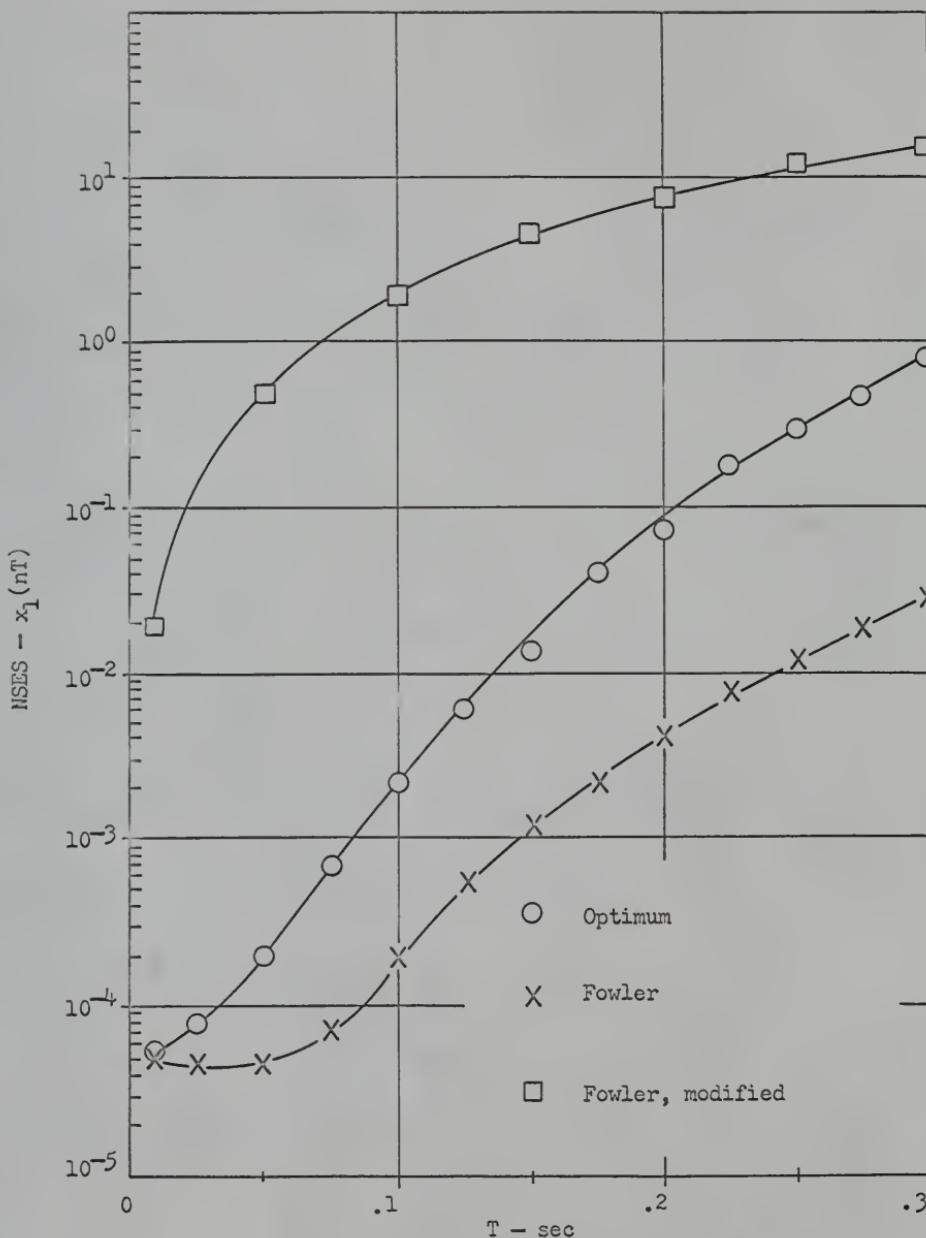
Figure 19. Error Criterion for x_1 of Nonlinear System with 10sin2t Input
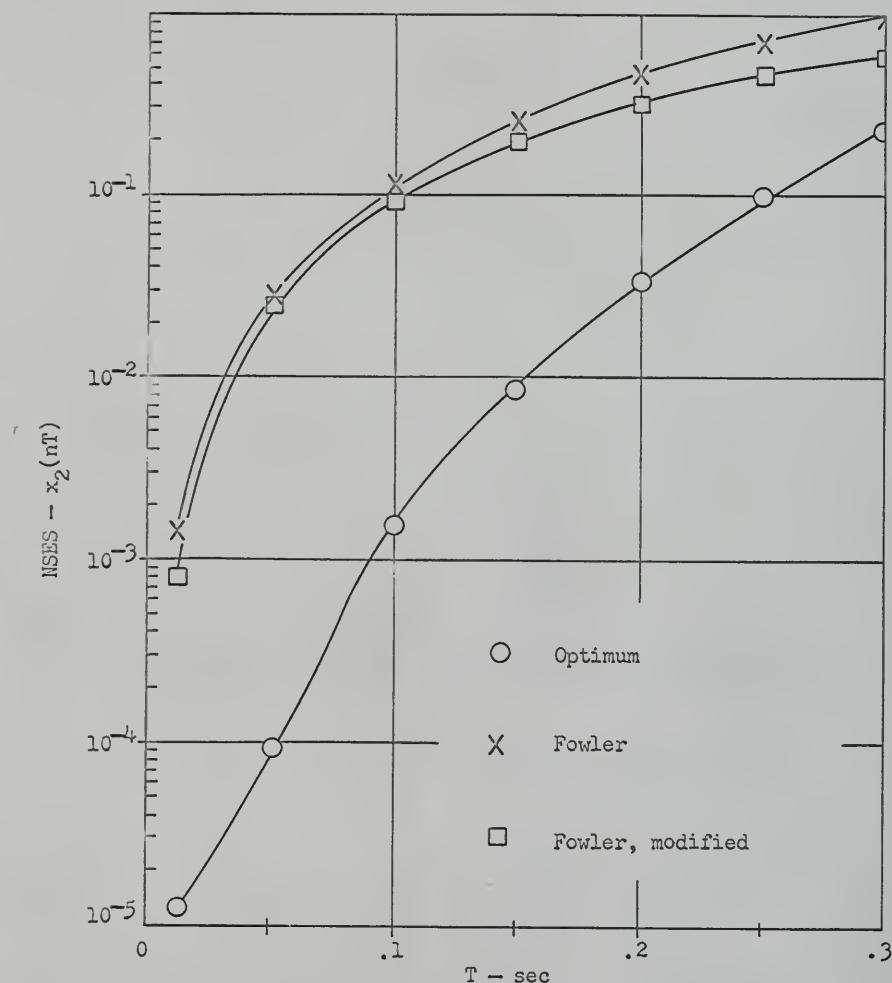
Figure 20. Error Criterion for x_2 of Nonlinear System
with 10sin2t Input
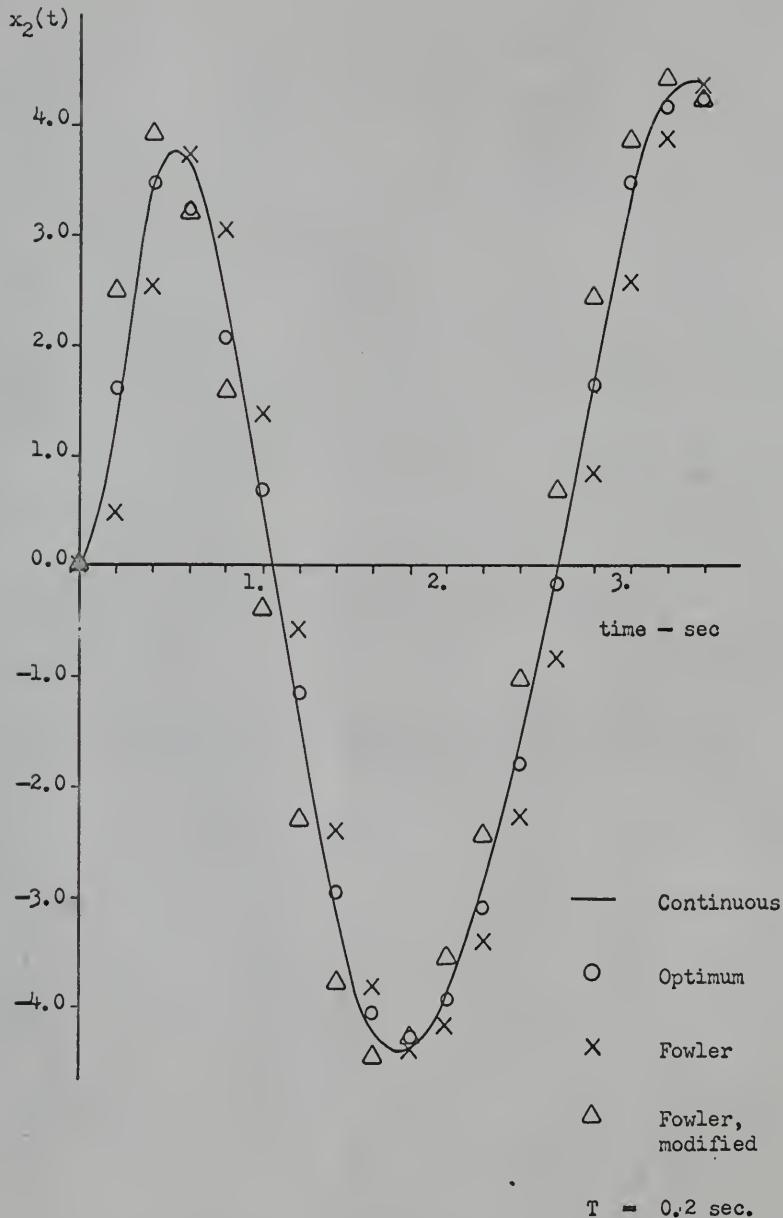
Figure 21. Nonlinear System $x_2(t)$ Response for 10sin2t Input
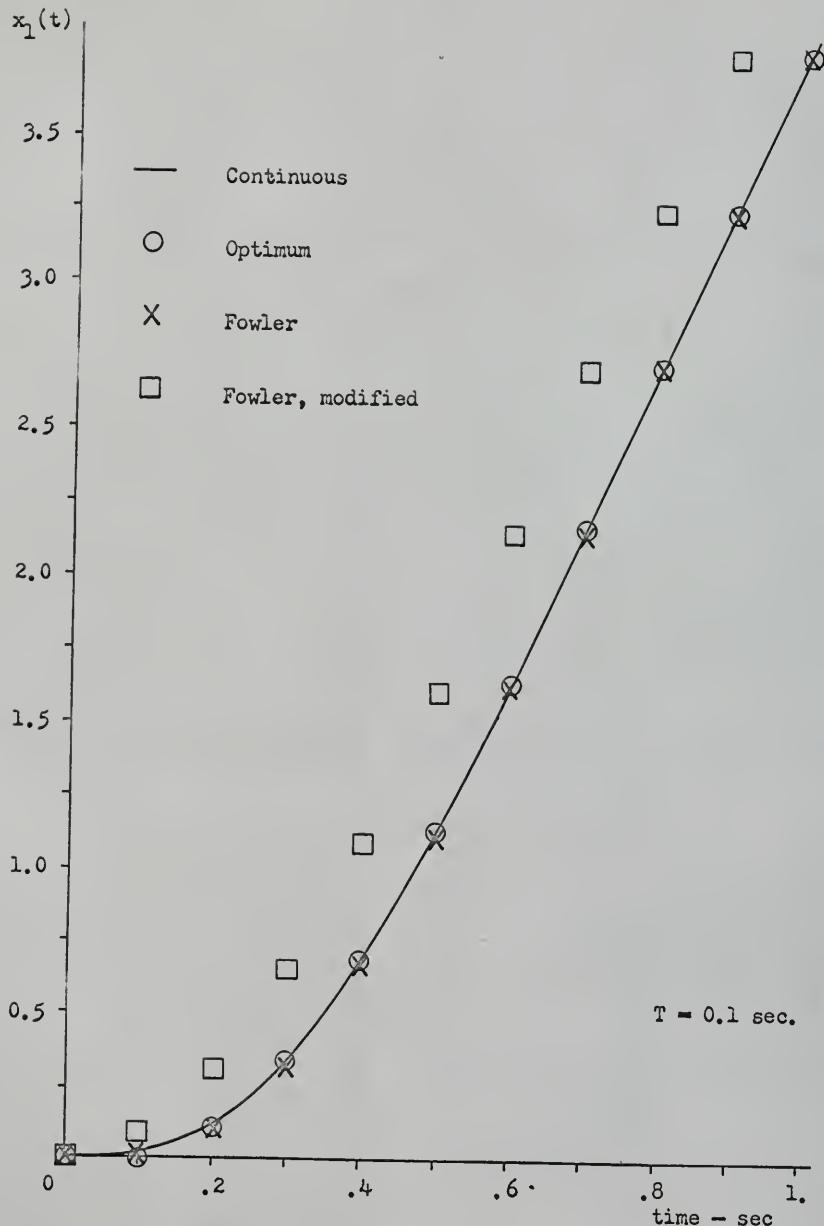
Figure 22. Response of the Nonlinear System to a Ramp Input

for what may be equally well termed a modified z—transform model.
Because of the close association of the Fowler method and z—transform
concepts, concentration on the former method was felt to be most
rewarding.

Uncertainty in the time of occurrence of input signals is a source
error in any discrete system or discrete simulation. Consideration of
this problem places additional limits on the permissible magnitude of
sampling interval. By making some assumptions regarding the statisti-
cal properties of the time of arrival of signals, it is possible to
determine bounds for the error associated with given types of input
signals and sampling interval size. Results regarding research on this
aspect of system discretization have been discussed by Sage and Melsa
[37].

<div align="center">Summary</div>

A comprehensive study of significant classical and modern tech-
niques for discrete modelling of differential systems has been pre-
sented. Discussion of the basic application procedures for each method
has been supported by extensive digital computer experimentation,
providing a basis for comparison of the different techniques. It has
been demonstrated that recently introduced methods for discretization
of differential systems permit accurate digital simulation of such
systems and yield performance at least equal to that of the classical
methods for linear systems simulations. For discretization of non-
linear systems, the Fowler method and optimum discrete approximation
method have been shown to lead to superior discrete models. The
specific nature of the examples investigated is recognized, as well as
the corresponding limitations impressed on possible conclusions re-
garding the experimental results. Such limitations are inherent in any

study of nonlinear and discrete systems and are accepted as normal
constraints on the discussion of results.

Due to the limitations of the timing facilities in the present
IBM 1401-709 data processing system, and the necessity for programming
economy, it has not been possible to establish absolute evidence of
real-time digital simulation. The coarse time increment available on
the digital machine only permits bounds to be set on the simulation
running time. Indications obtained from timing program segments for
simulations of different sampling interval size are that real-time
simulation is achieved by the optimum discrete model and, by impli-
cation, other techniques for the same order difference equation. For
the example considered here, the optimum discrete model permits imple-
mentation of a more rapid simulation since a larger sample interval
size may in some cases be employed for a given simulation error.

The experimental data presented for the discrete approximation
techniques considered give evidence of decided advantages for the
optimum discretization approach. The model derived via this method has
been shown to be of adequate and frequently superior simulation capa-
bility. In particular, the optimum discrete representation compares
favorably with the Fowler approximation. In addition to desirable
performance characteristics, the optimum discrete model is readily
developed for most systems through utilization of discrete forms availa-
ble in Table 2 and others which may be derived in the course of work
with the method and retained. For complex systems, the discretization
may be accomplished on a segmental basis, not requiring extensive
digital computer analysis as has been suggested for the Fowler method
formulation. For these reasons the optimum discrete approximation

method appears to be particularly promising for many applications.

Experimental results obtained in the course of this study indicate that further improvement is possible in the discrete model characterizations. The reduced sensitivity to sampling interval change of the Fowler model appears related to the parameter adjustment carried out in developing the discrete system. Similar adjustment of parameters in the discrete representations derived by other methods seems a promising avenue for improved modelling.

CHAPTER 3

IDENTIFICATION

The experimental study of discrete approximation techniques discussed in the last chapter revealed a possible avenue of approach to the improvement of discrete models. Adjustment of gain parameters in a model for improvement of the approximation is not a new concept, but the procedure normally used has been based on the intuition and experience of the experimenter. The Fowler method of discretization presents a first approach to an orderly attack on the problem. Efforts made in the areas of system identification and control have employed parameter identification techniques [38,39,40,41] which may be adapted for a new approach to the parameter adjustment in discrete models. In recent research on this aspect of discrete modelling, Sage [10,11] has introduced a formulation for the problem of parameter identification employing the methods of quasilinearization and differential approximation. Since these techniques have not been previously applied in this manner for the study of discrete systems, an intensive effort has been made to examine the problem formulation and experimentally study the properties of the identification procedures.

Quasilinearization

The method of quasilinearization for the resolution of boundary-

value problems arising in the solution of nonlinear differential
equations has been widely applied by Bellman in earlier referenced
works. A formulation of the method for two—point boundary—value
problems had been discussed by McGill and Kenneth [42] who refer to the
method, perhaps more properly, as the generalized Newton—Raphson tech—
nique. Application of this technique to difference equations appears
to have been lightly treated. Henrici [43] discusses this approach in
relation to a finite difference scheme for solution of a class of non—
linear boundary—value problems of second—order, and offers a proof for
convergence of the proposed scheme. A similar approach to solution
of two—point boundary—value problems via finite difference techniques
and use of quasilinearization has been presented by Sylvester and
Meyer [44].

Consider a system of nonlinear difference equations

$$\underline{x}(\overline{k+1}\ T) = \underline{f}(\underline{x}(kT),k), \qquad k \quad [L,K] \tag{3.1}$$

with boundary conditions

$$\left\langle \underline{c}_i\ (jT),\ \underline{x}(jT) \right\rangle = d_i(jT), \quad j = L,K$$

$$i = 1,2,\ \ldots\ldots,\ m/2, \tag{3.2}$$

where $\underline{x}$ and $\underline{c}$ are m—vectors, $\left\langle\ ,\ \right\rangle$ denotes the inner product, and
the period of observation of the solution vector is $t = LT$ to $t = KT$.
Utilizing the method of quasilinearization, an iterative procedure is
established for successively approximating the solution to Equation

(3.1) by solutions of the system of linear equations

$$\underline{x}^{q+1}(\overline{n+1}\ T) = \underline{f}(\underline{x}^q(nT),\ n) + J(\underline{x}^q(nT),n)\ [\underline{x}^{q+1}(nT) - \underline{x}^q(nT)], \qquad (3.3)$$

where J is the Jacobian matrix of partial derivatives having as its ij^{th} element the partial derivative $\partial f_i / \partial x^q_j$ and $\underline{x}^q$ indicates the solution at the qth iteration. Equation (3.3) may be stated as

$$\underline{x}^{q+1}(\overline{n+1}\ T) = A(nT)\underline{x}^{q+1}(nT) + \underline{B}(nT) \qquad (3.4)$$

where

$$A(nT) = J(\underline{x}^q(nT),n)\ ,$$

$$\underline{B}(nT) = \underline{f}(\ \underline{x}^q(nT),\ n) - J(\underline{x}^q(nT),\ n)\ \underline{x}^q(nT).$$

Since Equation (3.4) is linear in the $(q + 1)^{st}$ approximation, a solution may be determined by generating the homogeneous and particular solutions and imposing the boundary conditions of Equation (3.2). Let $\bigoplus^{q+1}(nT)$ be the fundamental matrix of

$$\bigoplus^{q+1}(n+1\ T) = A(nT)\ \bigoplus^{q+1}(nT), \qquad (3.5)$$

with

$$\bigoplus^{q+1}(L) = I, \text{ the m x m identity matrix.}$$

The particular solution $\underline{p}^{q+1}(nT)$ is generated by the equation

$$\underline{p}^{q+1}(\overline{n+1}\ T) = A(nT)\ \underline{p}^{q+1}(nT) + \underline{B}(nT), \qquad (3.6)$$

with

$$\underline{p}^{q+1}(L) = \underline{0} .$$

The solution for Equation (3.4) is now given by

$$\underline{x}^{q+1}(\overline{n+1}\ T) = \oiint^{q+1}(nT)\ \underline{v}^{q+1} + \underline{p}^{q+1}(nT) \tag{3.7}$$

with the constant vector $\underline{v}^{q+1}$ obtained from the boundary conditions by solving the equations

$$\left\langle \underline{c}_i(jT),\ \oiint^{q+1}(jT)\ \underline{v}^{q+1} + \underline{p}^{q+1}(jT) \right\rangle = d_i(jT), \tag{3.8}$$

$$j = L,K$$

$$i = 1,2,\ \ldots\ ,\ m/2 .$$

An initial or zeroth trajectory $\underline{x}^0$ may be generated by selecting initial conditions for the unknown elements of $\underline{x}(LT)$ and solving Equation (3.4) forward in time. In practice, Equation (3.7) is seldom used to obtain the $(q + 1)^{st}$ trajectory. Such an approach requires retention of the complete homogeneous and particular solutions trajectories with a corresponding requirement for computer memory storage. An approach having reduced memory storage requirements consists of retaining only $\oiint^{q+1}(LT)$, $\underline{p}^{q+1}(LT)$, $\oiint^{q+1}(KT)$, and $\underline{p}^{q+1}(KT)$ in memory until evaluation of $\underline{v}^{q+1}$ by Equations (3.8). The $(q + 1)^{st}$ trajectory $\underline{x}^{q+1}(nT)$ is then generated from Equation (3.4) with requisite initial condition vector elements obtained from $\underline{v}^{q+1}$. This trajectory is stored in computer memory as the final solution

or for evaluation of A(nT) and B(nT) for the next iteration.

Differential Approximation

The method of differential approximation introduced by Bellman [12,15,16,38] offers a direct approach to the problem of parameter estimation [45]. Because of the straightforward manner in which differential approximation is applied, it may serve as a convenient means for obtaining parameter estimates to initialize the quasilinearization procedure. This adds particular value to this procedure since the parameter values obtained from differential approximation are frequently quite poor and in need of refinement.

Consider an equation

$$\underline{y}(\overline{n+1}\ T) = \underline{f}(\underline{y}(nT),\underline{b}) , \quad n \ [L,K] \quad \text{where}$$

$y(nT)$ is an m-vector, and $\underline{b}$ is a parameter vector, an r-vector, to be determined so that $y(nT)$ closely approximates some vector $\underline{z}(nT)$. If some suitable parameter vector, $\underline{b}_o$, can be found so that

$$\underline{z}(\overline{n+1}\ T) = \underline{f}(\underline{z}(nT),\underline{b}_o) ,$$

this set of parameters with the initial conditions $\underline{z}(LT)$ will make $y(nT)$ identical with $\underline{z}(nT)$. Such a set of parameters may not exist; however, $\underline{b}$ may be determined to make

$$\underline{z}(\overline{n+1}\ T) - \underline{f}(\underline{z}(nT),\ \underline{b})$$

as near zero as possible. The set of parameters may then be chosen so that

$$\sum_{n=L}^{K} \left\| z(n+1\ T) - f(z(nT),\ b) \right\|^{2}_{R(nT)} \qquad (3.9)$$

is minimized with respect to $\underline{b}$, where $R(kT)$ is an m x m positive
semi-definite weighting matrix. The minimization may be accomplished
by equating to zero the partial derivatives of Equation (3.9) with
respect to the components of b, yielding equations

$$\sum_{n=1}^{K} \ [\ \nabla_{b}\underline{f}{\,'}\cdot(\underline{z}(nT),\underline{b})]\ [\underline{z}(\overline{n+1}\ T) - \underline{f}(\underline{z}(nT),\underline{b})] = \underline{0} \qquad (3.10)$$

where $\nabla_{b}f' = [\ \partial f_{j}\ /\ \partial b_{i}\ ]$, $R = I$, and $\underline{0}$ is the r dimensional
null vector. Solution of these r equations in the components of $\underline{b}$
produces the parameter estimates.

Formulation of an Approach to Discrete System Parameter Identification

It has been noted that the dependence of discrete approximation
error on sampling interval size may be reduced by modification of the
approximation in relation to changes in the sampling interval. A
similar dependence of approximation error on input signals occurs
in the simulation of nonlinear systems, and it appears that improve-
ment in the approximation error may also be achieved for this case by
suitable model adjustment. Having a discrete model, represented by
state vector $\underline{y}_{d}(nT)$, for a continuous system with state vector $\underline{y}_{a}(t)$,
it is desired to determine a parameter vector, $\underline{b}$, for the discrete

system so that $\underline{y}_d(nT) \rightarrow \underline{y}_a(nT)$ for a given input signal and sampling interval T. Knowing the continuous system response for a given input, the parameters are to be adjusted to minimize the sum of error squared between the continuous system state and the discrete system state,

$$J = \frac{1}{2} \sum_{k=0}^{N-1} \left\| \underline{y}_a(kT) - \underline{y}_d(kT) \right\|_{R(kT)}^2 \tag{3.11}$$

subject to the constraints

$$\underline{y}_d(\overline{n+1}\ T) = \underline{f}(\underline{y}_d(nT),\ \underline{b})\ ,\quad n \in [\,0,N\,] \tag{3.12}$$

$$\underline{b}(\overline{n+1}\ T) = \underline{b}(nT), \tag{3.13}$$

and

$$\underline{y}_d(0) = \underline{y}_a(0)\ , \tag{3.14}$$

where $\underline{y}_d(nT)$ and $\underline{y}_a(nT)$ are m-vectors, $\underline{b}(nT)$ is an r-vector, and R is an m x m positive semi-definite weighting matrix.

The minimization of J is accomplished via variational calculus procedures, utilizing a Lagrange multiplier formulation for discrete systems [46,47] . Adjoining the system constraints to J yields an augmented criterion

$$J^* = \sum_{k=0}^{N-1} \frac{1}{2} \left\| \underline{y}_a(kT) - \underline{y}_d(kT) \right\|_R^2$$

$$+ \mu'(\overline{k+1}\ T)\ [y_d(\overline{k+1}\ T) - \underline{f}(\underline{y}_d(kT),\ \underline{b})]$$

$$+ \beta'(\overline{k+1}\ T)\ [\underline{b}(\overline{k+1}\ T) - \underline{b}(kT)]\quad, \tag{3.15}$$

where the argument of $R(kT)$ has been omitted for convenience in writing and where μ' denotes the transpose of the vector μ. The resulting equations for the adjoint variables become

$$\mu(nT) = [\nabla_y \underline{f}'\ (\underline{y}\ (nT),\underline{b})]\ \ \mu(\overline{n+1}\ T) + R(nT)\ [\ y_a(nT) - \underline{y}_d\ (nT)], \tag{3.16}$$

and

$$\beta(nT) = \beta(\overline{n+1}\ T) + [\nabla_b \underline{f}'(\underline{y}_d(nT),\underline{b})]\ \ \mu(\overline{n+1}\ T) \tag{3.17}$$

with boundary conditions

$$\mu(\overline{N-1}\ T) = \beta(\overline{N-1}\ T) = \beta(0) = \underline{0}\ . \tag{3.18}$$

Equations (3.12), (3.13), (3.16), and (3.17) together with the boundary conditions pose a two-point boundary-value problem which may be resolved by the quasilinearization approach discussed above. The difference equations are transformed into a new family of variables by defining a $2(m + r)$ dimensional vector

$$\underline{x}'(nT) = [\ \underline{y}'\ \ (nT),\ \underline{b}',\ \mu'(nT),\ \beta'(nT)]\ . \tag{3.19}$$

This combines Equations (3.12), (3.13), (3.16), and (3.17), and may now be written as

$$\underline{x}(\overline{n+1}\ T)\ =\ \underline{g}(\underline{x}(nT))\ , \tag{3.20}$$

with boundary conditions from Equations (3.14) and (3.18)

$$\left\langle \underline{c}_i(\ jT),\ \underline{x}(jT) \right\rangle \quad =\ d_i(jT), \qquad j = 0, N-1$$
$$i = 1, 2, \ldots, (m+r)\ .$$
$$\tag{3.21}$$

An initial estimate of the parameter vector $\underline{b}$ permits Equation (3.20) to be solved, yielding an initial trajectory $\underline{x}^o(nT)$. The $(q+1)$st approximation to the solution is given by Equation (3.3) which here becomes

$$\underline{x}(\overline{n+1}\ T)\ =\ [\nabla_x g'(\underline{x}^q(nT))\ ]'\ [\ \underline{x}^{q+1}(nT)\ -\ \underline{x}^q(nT)]\ +\ g(\underline{x}^q(nT),$$
$$\tag{3.22}$$

where $\left[\nabla_x g'\right]'$ is the Jacobian matrix earlier defined. The successive approximations to $\underline{x}(nT)$ and estimates for the parameter vector $\underline{b}$ are obtained by the procedure of Equations (3.5) through (3.8).

The computational load required by the implied matrix inversion of Equation (3.8) may be alleviated by reducing the order of the coefficient matrix from the a priori knowledge of the boundary conditions

$$x_i(0)\ =\ y_{ai}(0)\ ,\quad i\ = 1, 2,\ \ldots,\ m \tag{3.23a}$$

$$x_i(0)\ =\ 0, \qquad i\ = 2m + r + 1,\ \ldots,\ 2(m+r) \tag{3.23b}$$

$$x_i(\overline{N-1}\ T) = 0,\ i = m + r + 1,\ m + r + 2,\ \ldots,$$

$$2(m + r). \qquad (3.23c)$$

The necessary values for the b_i and $u_i(0)$ may be obtained by inverting only an $m + r$ square matrix consisting of ij^{th} terms of $\Phi(\overline{N-1}\ T)$ where i ranges from $m + r + 1$ to $2(m + r)$ and j ranges from $m + 1$ to $m + r$. The corresponding constant vector consists of the final values of the particular solution vector, $-p_i(\overline{N-1}\ T)$, for $i = m + r + 1$ to $i = 2(m + r)$. This modification of the procedure requires that only the final values of the homogeneous and particular solutions be retained in the computer memory. Termination of the iteration process is caused by satisfaction of some desired criterion such as a specified rate of change of parameter values.

Differential approximation may be readily applied to this problem for parameter identification or initialization of the quasilineariza-tion procedure. This is achieved by applying the conditions of Equation (3.10) to Equation (3.12) with the result

$$\sum_{n=0}^{N-1} [\nabla_b \underline{f}'(\underline{y}_a(nT),\ \underline{b})]\ [\underline{y}_a(\overline{n+1}\ T) - \underline{f}(\underline{y}_a(nT),\underline{b})] = \underline{0}. \qquad (3.24)$$

Experimental Study of Parameter Identification

Digital computer experiments designed to test the effectiveness of the identification procedures formulated above have been organized in relation to the work in the last chapter. Since studies were made

there of the dependence of discrete approximation error on sampling interval size, the data from that study supplies a convenient basis for comparison in an attempt to improve discrete model characteristics through parameter adjustment. The first experiment undertaken is an application of the quasilinearization procedure to the nonlinear system example of Chapter 2. The system configuration is repeated in Figure 23 for convenient reference, where the system state variables are now termed $y_{a1}(t)$ and $y_{a2}(t)$. The parameter identification procedure is implemented for the optimum discrete approximation to the continuous system. The difference equations for this approximation
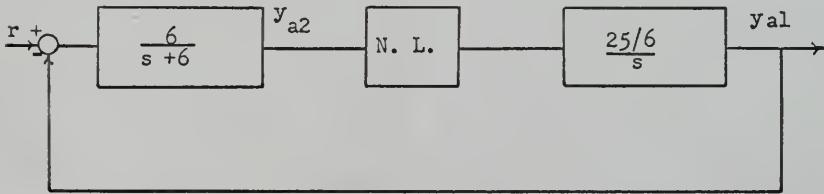

Figure 23. Nonlinear System for Parameter Identification

were stated in Equations (2.56) and are restated here, calling the states $y_{d1}(nT)$ and $y_{d2}(nT)$ and including in the expressions parameters b_1 and b_2 which are to be identified.

$$y_{d1}(\overline{n+1}\ T) = y_{d1}(nT) + \frac{25}{12}\ T\ b_2(nT)\ [4h(nT) - 3h(\overline{n-1}\ T) + h(\overline{n-2}\ T)]$$

$$y_{d2}(\overline{n+1}\ T) = ay_{d2}(nT) + b_2(nT)\ [d_1 e(\overline{n+1}\ T) + d_2 e(nT)]$$

$$h(nT) = y_{d2}(nT) + .01\ y^3_{d2}(nT)$$

$$e(nT) = r(nT) - y_{d1}(nT)$$

$$d_1 = (6T - 1 + a)\ /\ 6T$$

$$d_2 = (1 - a - 6Ta)\ /\ 6T$$

$$a = e^{-6T}. \tag{3.25}$$

Development of the equations for the adjoint variables gives rise
to a problem in the formulation. Desiring to obtain an expression
for $\mu(\overline{n+1}\ T)$ from Equation (3.16), the matrix $\nabla_y \underline{f}'(\underline{y}_d(nT), \underline{b})$ must
be inverted. When Equations (3.25) are placed in true discrete
state variable form, as in Equation (3.12), so that all terms on the
right hand side are of argument nT, the matrix resulting from the
gradient operation, $\nabla_y \underline{f}'(\underline{y}_d(nT), \underline{b})$, is "almost" singular. Such
a poorly conditioned matrix produces equations of correspondingly
poor stability. When implemented on the computer, the effects of
rounding numbers within the machine are apparently enough to cause
certain singularity of the matrix.

The problem of instability of the adjoint equations may be
circumvented, for values of the b_i near the optimum, by considering
variables with delayed arguments in Equation (3.25) as coefficient
terms, and obtaining the gradient matrix $\nabla_y \underline{f}'$ by taking

derivatives with respect to $y_{d1}(nT)$ and $y_{d2}(nT)$. For choices of
the b_i which are not sufficiently near optimum, instability appears
to be generally unavoidable. The approximate gradient matrix
$\nabla_y f'$ now appears as

$$
\begin{bmatrix}
1 & b_2(1-a) \\
\\
s & \alpha - d_1 b_2 s
\end{bmatrix}
\tag{3.26}
$$

$$
s = \frac{25}{3} T\, b_1\, [1 + .03\, y^2_{d2}(nT)] \quad ,
$$

where a and d_1 are as defined in Equations (3.25). Since detailing
of all the matrices involved in developing the examples would appear
to add little to the presentation, they will not be generally shown.
A concise statement for $\mu(\overline{n+1}\ T)$ and $\beta(\overline{n+1}\ T)$ is obtained from
Equations (3.16) and (3.17) as

$$
\mu(\overline{n+1}\ T) = [\ \nabla_y f'\ (y_d(nT),\ b)\ ]^{-1} [\mu(nT) + y_d(nT_- - y_a(nT)]\ ,
\tag{3.27}
$$

and

$$
\beta(\overline{n+1}\ T) = \beta(nT) - [\ \nabla_b f'(y_d(nT),\ b)]\ \mu(\overline{n+1}\ T),
\tag{3.28}
$$

where R has been made a 2 x 2 identity matrix. A new state vector
is obtained in the form of Equation (3.20) by adjoining Equations (3.27)

and (3.28) to Equations (3.25) together with the vector $[b_1, b_2]'$.
The boundary conditions are given by

$$y_d(0) = y_a(0) = \underline{0} \; ,$$

and

$$\underline{\mu}(\overline{N-1}\;T) = \underline{\beta}(\overline{N-1}\;T) = \underline{\beta}(0) = \underline{0} \; . \tag{3.29}$$

The linearized equations for the iteration procedure now require
derivation of the Jacobian matrix $[\nabla_x \underline{g}']'$ which for this example
is an 8 x 8 matrix whose ij^{th} element is $\partial g_i / \partial x_j$. The system of
linear, time-varying equations are again stated as

$$\underline{x}^{q+1}(\overline{n+1}\;T) = [\nabla_x \underline{g}'(\underline{x}^q(nT)]' [\underline{x}^{q+1}(nT) - \underline{x}^q(nT)] + \underline{g}(\underline{x}^q(nT)).$$

The homogeneous and particular solutions of the linearized equations
are now sought. The homogeneous solution is obtained from

$$\underline{\Phi}^{q+1}(\overline{N-1}\;T) = \prod_{k=0}^{N-2} [\nabla_x \underline{g}'(\underline{x}(nT))]' \quad \underline{\Phi}^{q+1}(0) \tag{3.30}$$

where $\underline{\Phi}^{q+1}(0) = I$, an 8 x 8 identity matrix. It is noted
here that only $\underline{\Phi}^{q+1}(\overline{N-1}\;T)$ and $p^{q+1}(\overline{N-1}\;T)$ need be retained for
evaluation of the initial condition vector $\underline{v}^{q+1}$. The particular
solution vector is given by

$$p^{q+1}(\overline{N-1}\;T) = \sum_{j=0}^{N-2} H(N-1, j+1) \; \underline{B}(jT), \tag{3.31}$$

where

$$H(N-1, j+1) = \prod_{k = j+1}^{N-2} [\nabla_x g'(\underline{x}(kT))] \qquad ,$$

and

$$\underline{B}(jT) = g(\underline{x}^q(jT)) - [\nabla_x g'(\underline{x}(jT))]' \; \underline{x}^q(jT),$$

since $\underline{p}^{q+1}(0) = \underline{0}$. Flow diagrams for the computer programs to accomplish these operations are presented in Appendix IV. From the boundary conditions of Equations (3.29), it is recognized that $v_i^{q+1} = 0$ for i = 1,2,7,8. It is then necessary to evaluate only the remaining four initial conditions at each iteration.

For the example system with a 10 unit step input, the gain parameters b_1 and b_2 were determined for sampling intervals of .01 to 0.3 second for an observation time of 5 seconds. The most rapid convergence of the process was achieved by starting the experiments at the low sample period where the b_i are approximately unity. For successive programs with increased sample periods, the parameter values obtained from the previous experiment were employed as starting estimates. The iteration process for each experiment was terminated on a successful convergence by halting the iterative procedure when parameter values changed less than 10^{-4} between successive iterations. Since no "true" value of parameter was known, the procedure was forced to run for a minimum number of iterations, usually three to five. Table 5 illustrates the convergence of the gain parameters for

the present example for the case where the sampling interval is
.1 second. The numbers shown have been rounded to six decimal
places and in the actual result do change less than 10^{-6} between
iterations 3 and 4.

Table 5

Gain Parameter Convergence

Iteration	b_1	b_2
0	1.000000	1.000000
1	.846688	1.035062
2	.820133	1.041022
3	.819920	1.040923
4	.819920	1.040923

Figure 24 illustrates the parameter adjustment obtained via
the quasilinearization procedure as the sampling interval was changed.
The related effect on the approximation error dependence on sampling
interval is shown by the data of Figures 25 and 26 which display the
value of the NSES criteria for x_1 and x_2 . In these figures the data
from the experiments of Chapter 2 are repeated to demonstrate the
effect of gain parameter change. It is apparent that there has been
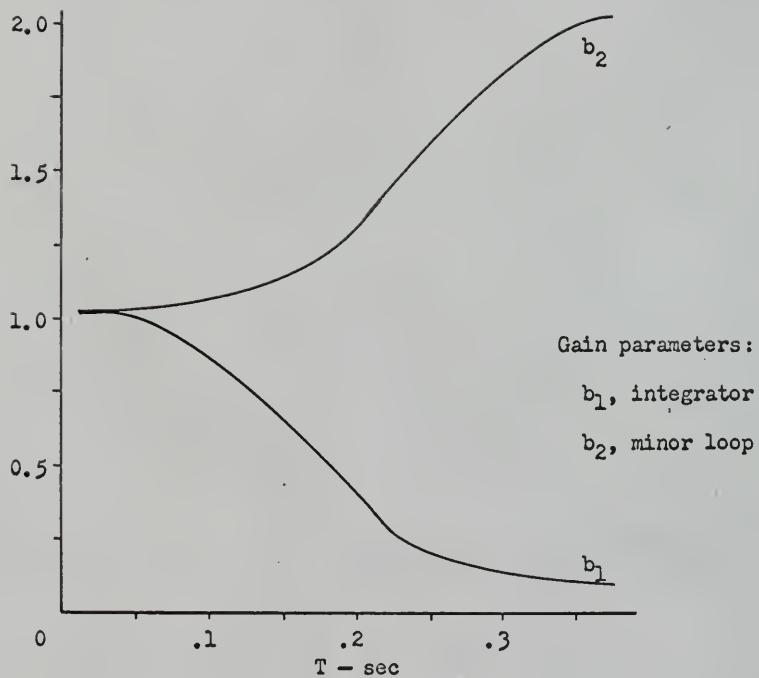
Figure 24. Discrete Model Gain Adjustment via Quasilinearization
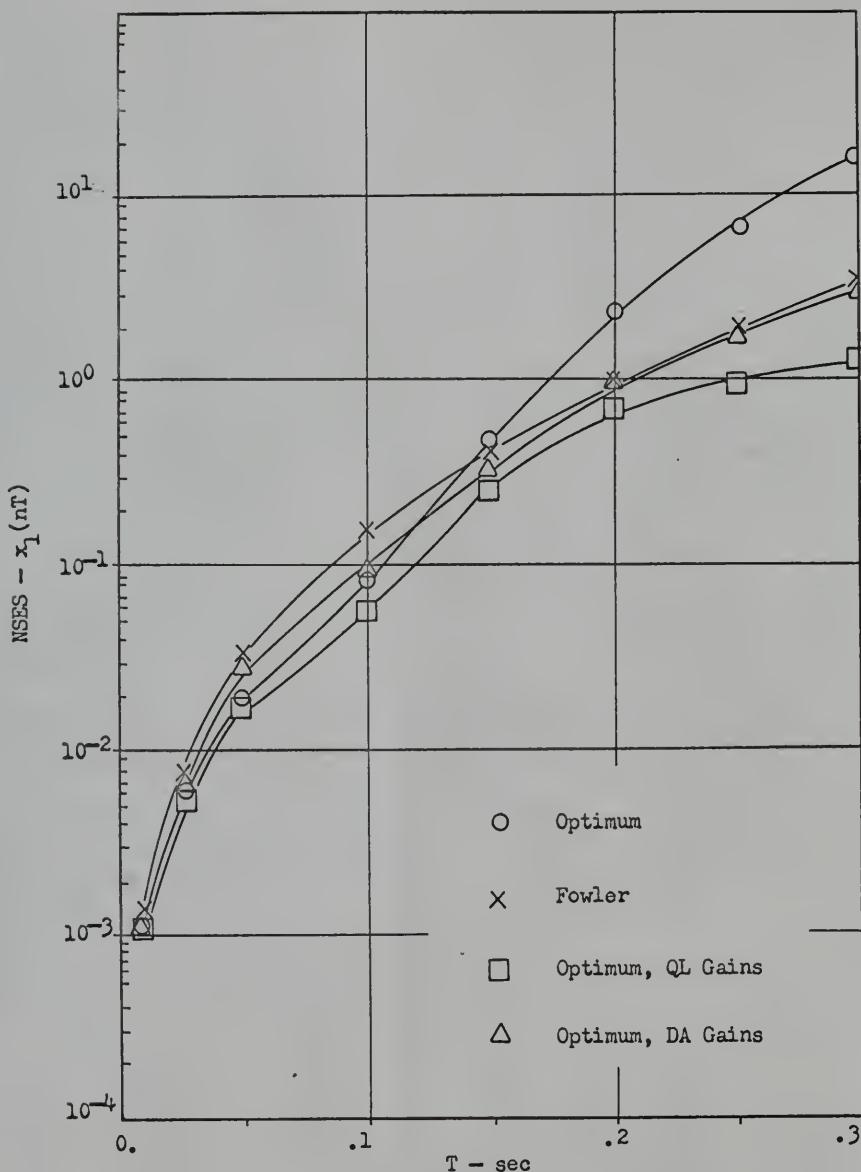
Figure 25. Error Criterion for x_1 with Identified Gain Parameters, Step Input
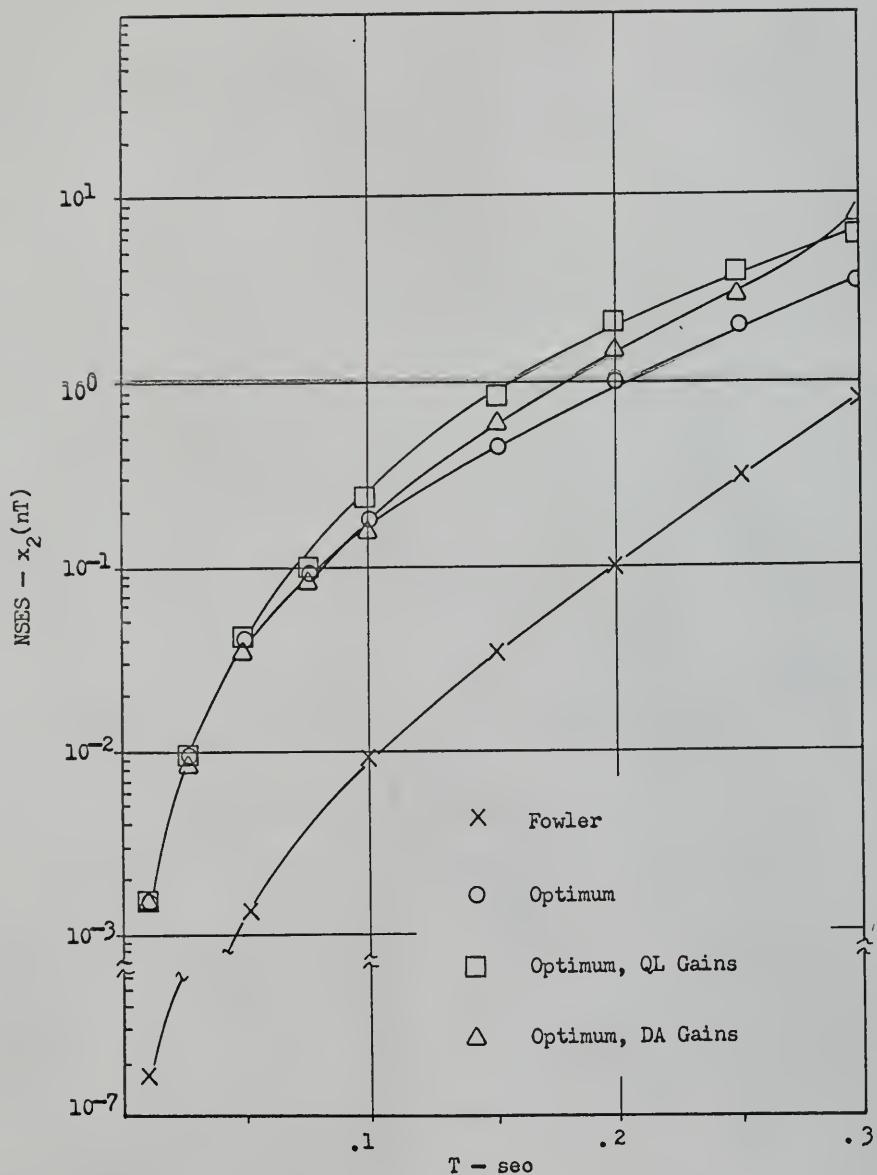
Figure 26. Error Criterion for x_2 with Identified
Gain Parameters, Step Input

some achievement in reducing the error in x_1, but at the expense of x_2. It should be recalled that the weighting matrix in the criterion of Equation (3.11) was taken to be an identity matrix for this example. Some further experimentation with the form of the weighting matrix appears to hold promise as a means of tailoring model response for desired error distribution. The step response of the system is shown in Figure 27 further illustrating the effect of parameter adjustment.

For the same conditions of input signal and sampling interval, the differential approximation method was used to determine the parameter adjustment. The expression of Equation (3.24) was formed employing for $y_a(nT)$ values of system response obtained via the Runge-Kutta integration method. The effect of these gain settings on the simulation error is shown in Figures 25 and 26. Like the quasilinearized gains these have negligible effect on the error at small sampling intervals where the error is very low, but there is improvement at the larger sampling intervals. The error reduction achieved by differential approximation is not so beneficial as seen in the data for x_1 but the effect on x_2 is somewhat less damaging. Here again uniform weighting was employed in the problem formulation. The gain parameter variation with sampling interval as determined via differential approximation is shown in Figure 28.

A second experiment utilizing the method of quasilinearization was performed to study the manner in which the gain parameter values are affected by changing input magnitude of the nonlinear system.
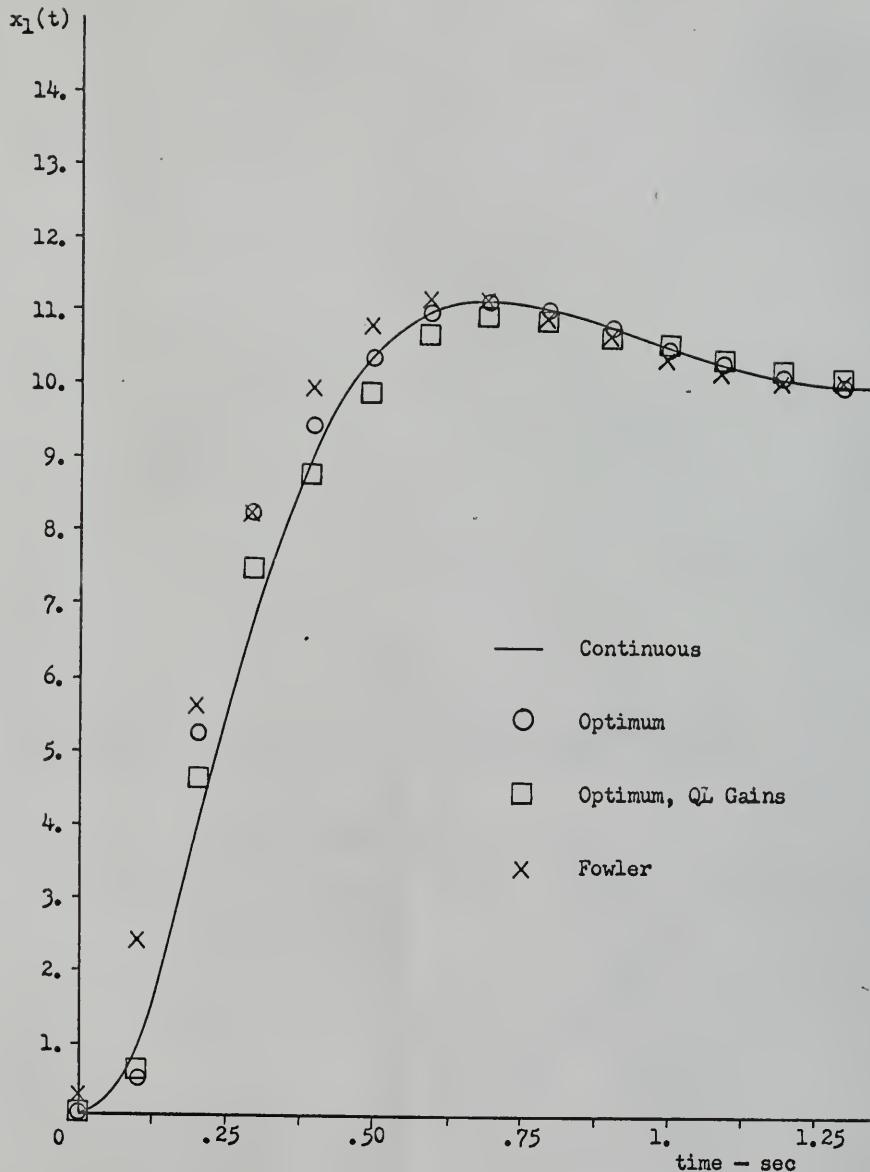
Figure 27. Effect of Gain Parameter Change on Step
Response of Nonlinear System
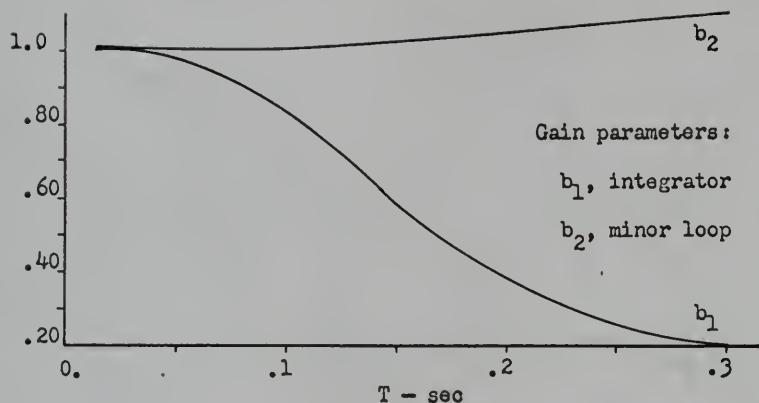
Figure 28. Gain Adjustment via Differential Approximation
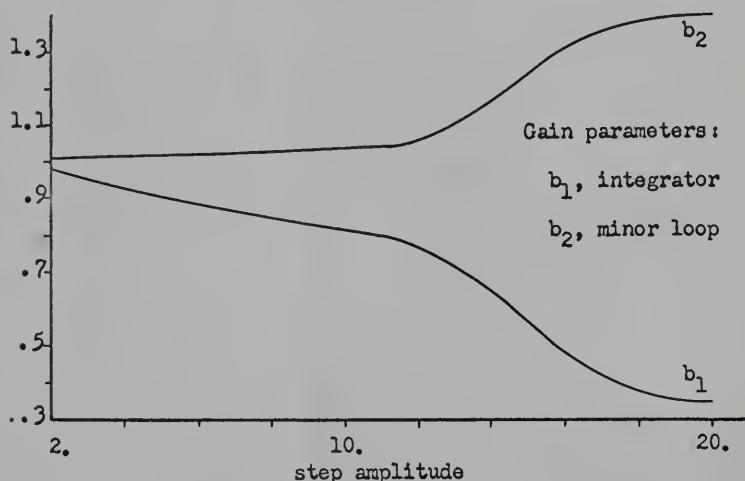
Figure 29. Gain Adjustment for Input Step Change

With a sampling interval size of .1 second, the input step function
amplitude was varied from .5 unit to 20. units. The gain parameter
values obtained with the quasilinearization method are shown in Figure
29. The principal result of the gain change to larger magnitude
input step functions is a reduction in the initial overshoot in the
response of the optimum discrete simulation. This is reflected in
the NSES criterion for the output state x_1 for which the value with
with 20. unit step function input is reduced from 1.659 to 0.904.
Here again the gain adjustment is of no benefit to x_2 , for which the
criterion value is slightly increased from unity gain magnitude of
0.808 to an adjusted magnitude of 0.821.

Experimentation performed with a sine wave input to the system
yielded less promising results than for the step input and made
apparent the convergence problems which constitute a major weakness
in the quasilinearization procedure. Figure 30 displays the data
for the NSES criterion for x_1 with an input of 10sin2t. The data
for the quasilinearized gains are not extended beyond a sampling
interval of .15 second due to the uncertainty introduced into the
results by the poor convergence properties of the method as the
sampling interval is increased. For the sampling interval range in
which data points are given, some problems occur but reliable
convergence appeared to be obtained after some experimentation with
initial values. The convergence problem is illustrated by the data
of Table 6. Displayed are the gain parameter values obtained from
different initial estimates for a sampling interval of 0.2 second.
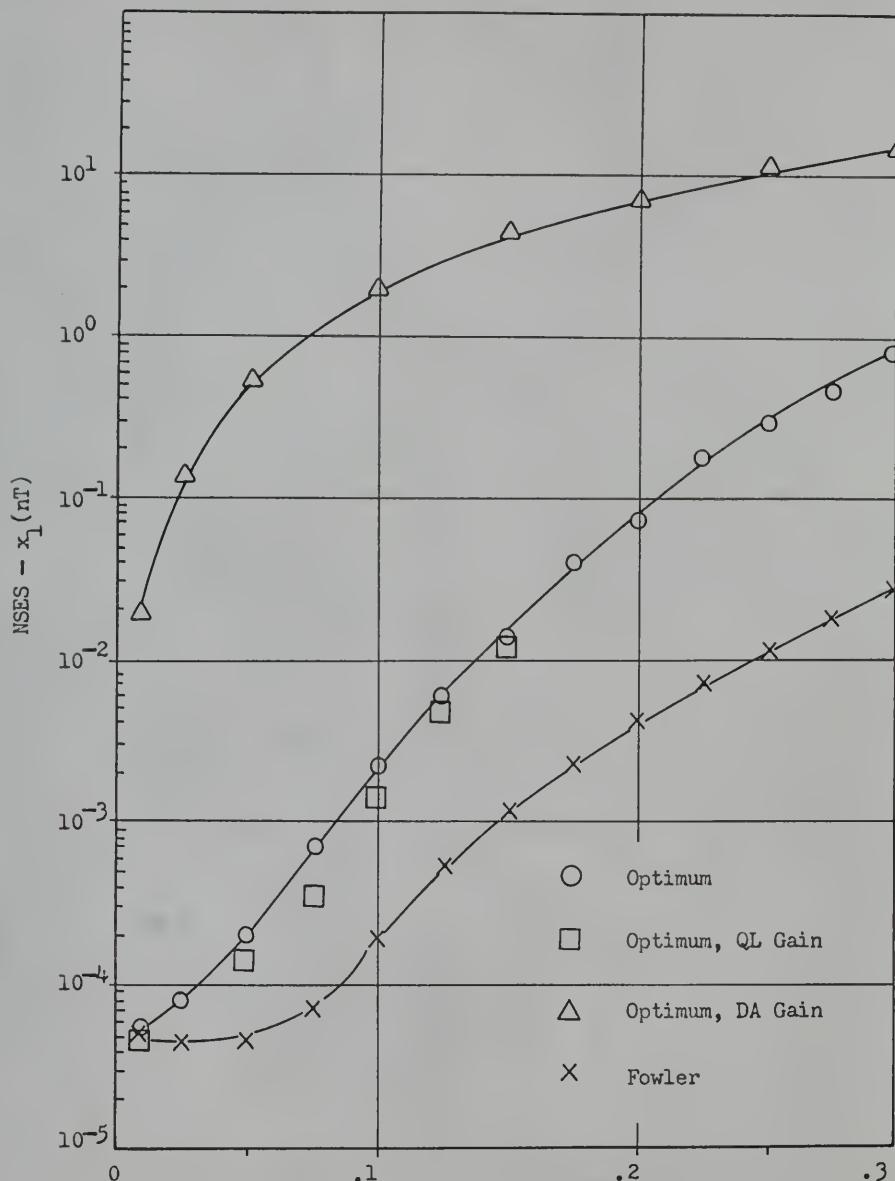
Figure 30. Error Criterion for x_1 with Sine Input and Identified Gains

Table 6

Convergence of Quasilinearization
for Nonlinear System with 10sin2t Input and T = 0.2 Second

Iteration	First Run		Second Run	
	b_1	b_2	b_1	b_2
0	.700000	1.200000	.945788	1.022408
1	.824744	1.866403	.900486	1.036902
2	.781149	1.183690	.905106	1.035109
3	.783074	1.181670	.904995	1.035158
4	.782964	1.181708	.904996	1.035158
5	.782963	1.181708	.904996	1.035158

NSES (x_1)	.1033	.0370
NSES (x_2)	.2456	.0223

and the accompanying values of the NSES criteria for the state
variables. It is noted that convergence is obtained for both cases
with quite different gains resulting and with correspondingly quite
different criteria values of which none improves the unity gain case.

Identification of the gain parameters via differential
approximation was also attempted for the case of the sine wave input.
These NSES criterion values for x_1 are also shown in Figure 30. The
results obtained reveal the complete failure of the approach in this
case. The gain parameter values obtained are so far from the expected
neighborhood of optimum values that they could not serve as starting
values for quasilinearization with convergence resulting even at
small sample intervals.

Application of the quasilinearization and differential
approximation methods to identification of time-varying system
parameters is a matter of interest in attempting to obtain a useful
discrete approximation to physical systems. As an example of such
a case, the nonlinear system studied earlier was modified by
replacing the nonlinearity with a time-varying gain. The resulting
system is shown in Figure 31, for which system it is desired to
evaluate the parameter b in the gain having records of the state
variable trajectories for a step function input. The discrete
system state equations are given by Equations (3.25) where the b_1
and b_2 there are made unity and $h(nT)$ becomes $b\sin 2nT$.
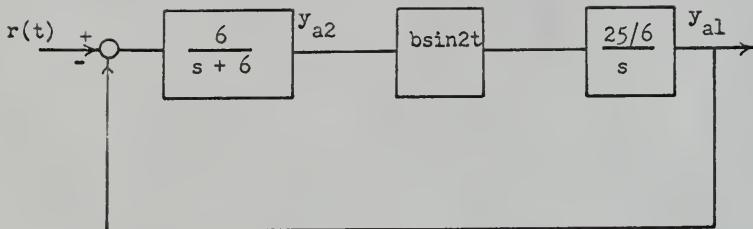
Figure 31. Time-Varying System for Identification

The system of equations to which the quasilinearization is applied is now a sixth-order system with boundary conditions

$$x_1(0) \;=\; x_2(0) \;=\; y_{a1}(0) \;=\; y_{a2}(0) \;=\; 0,$$

and

$$x_5(\overline{N-1}\ T) \;=\; x_6(\overline{N-1}\ T) \;=\; x_1(0) \;=\; 0$$

This experiment was performed for a sampling interval of .05 second and an input step of magnitude 2. The true value of the gain parameter was 1.25. Table 7 illustrates the convergence of the process for an initial parameter estimate of 1.30.

It is noted that the cost criterion NSES for the output is in this case .0251 in contrast to the value .1749 when the true gain

amplitude is employed in the discrete model. Implementation of the
differential approximation method employing Equation (3.24) yielded
a parameter value of .1305.

Table 7

Convergence of the Quasilinearization
Process for the Nonstationary System Example

Iteration	Parameter
0	1.300000
1	1.299847
2	1.280312
3	1.273894
4	1.273173
5	1.273170

The examples reported here employed the same basic control
program for the quasilinearization implementation on the digital
computer. Each example however required derivation of the system
difference equations, and notably tedious, derivation of the Jacobian
matrix. A minor attempt to mechanize the Jacobian determination
through use of basic difference techniques did not yield promising
results in terms of convergence of the process. Since convergence
difficulties were encountered with the program utilizing exact

expressions for the Jacobian, it appears that in general the errors involved in the difference calculations for the approximate Jacobian will not enhance the convergence properties of the process. It is observed that the adjoint variables quickly change from monotone to oscillatory functions with small change in the estimated parameters. Such behaviour places stringent requirements on an approximate method for taking derivatives.

For the nonlinear example in which two gain parameters are estimated, it is apparent that any number of combinations of parameter values will yield the same total cost or the same output response approximation error. The input signal amplitude for which the nonlinearity just begins to exert influence produces a particularly critical situation with ambiguous values for the parameters resulting from different choices of initial parameter values. The values obtained in this case yield approximately the same total cost and the same cost for state x_1 but with different costs for x_2 . This ambiguity in possible values for the parameters may be a contributing factor in the convergence problem and employment of some suitably defined weighting factor may prove of value.

Summary

A discrete formulation of the quasilinearization and differential approximation procedure for discrete system identification has been developed and made the subject of a program of digital experiments. While no analytical basis is established for the discrete version of these procedures, the experimental results parallel those for

continuous time systems in several respects and provide a computational justification for the formulation. The requirement for a bounded Jacobian matrix to assure convergence of the quasilinearization method in the continuous time domain has been observed also to hold in the discrete-time domain. This condition arose in initiating the discrete quasilinearization formulation where it became necessary to develop an approximate gradient matrix to obtain a Jacobian matrix with finite elements over the time interval of interest. The convergence properties of the discrete quasilinearization procedure parallel those in the continuous time domain in that the region in which initial unknown parameter values may occur and give convergence is very critical. The discrete formulation offers an advantage in this regard for the adjoint variables. Since the difference equations of the system may be arranged to run backward in time, it becomes necessary to estimate only the initial gain parameter values. The initial values of the adjoint variables for the zeroth iteration trajectories are obtained by backward difference operations. This approach has proved quite valuable in obtaining convergence of the identification procedure with a minimum number of computer runs. Discretization of the differential approximation method follows a straightforward translation from the continuous time domain and yields results of similar quality.

It has been shown that discrete quasilinearization method is of value in the identification of discrete systems. Improvement in discrete model performance has been achieved through application of the method to parameter identification. While application has

been made to only the optimum discrete approximation, the method
is applicable to any discrete system identification problem.
An expanded utilization of the approach appears possible through
a segmental approximation technique [38] for nonstationary systems.

Problems areas encountered in the experimentation with the
identification scheme require further consideration. Additional
experimentation concentrating on the convergence properties of the
method as related to input signal characteristics could yield
information for an improved formulation. Analytical work with
regard to definition of necessary conditions for convergence and
establishment of useful relations for selection of initial parameter
estimates is needed.

The gain parameters identified for the nonstationary system
example revealed that an optimum discrete parameter value may not
be the true continuous system parameter value. A similar result
is noted in the nonlinear system example. For that case, the pulse
transfer function gains do not become unity for small sample
intervals although the values approach magnitudes of unity. Both
these results indicate that the best discrete model is not obtained
by a direct replacement of continuous system gains or nonlinearities.
Identification of discrete model parameters for simple approximations
of such terms may produce an improved model.

Detailed statements of the difference equations were not given
for the examples solved in the experimental study. The high-order
systems resulting from the problem formulation creates an expanding
set of expressions not conveniently written in the available space.

It is observed that the second-order system with two unknown parameters was transformed into an eighth-order system for solution by the identification method. For realistic engineering problems of some complexity, it appears unfeasible to handle in the present manner a complete system with many parameters to be identified. Many applications should permit segmenting a system by loops or groups of elements which may be identified on a local basis. Some final modification of a few selected parameters might then be attempted.

CHAPTER 4
CONCLUSIONS

An investigation of classical and modern techniques for discrete representation of continuous dynamic systems has been conducted and detailed comparisons made regarding the effectiveness of these methods for digitally simulating linear and nonlinear systems. Procedures for application of discrete modelling methods to continuous system transfer functions are given and illustrated by example formulations. Advantages of more recently developed discretization techniques are shown and possible avenues of improvement in discrete modelling methods discussed.

Formulation of a parameter identification procedure for improved approximations has been developed, leading to a two-point boundary-value problem which was resolved via the method of quasilinearization. A second approach to parameter identification utilizing the method of differential approximation has also been presented. A discrete formulation of these procedures was developed for the digital computer. A series of experiments performed on the computer has demonstrated the effectiveness of these methods and has also revealed areas of interest for future work.

Examples studied herein indicate that certain discrete model parameters are more significantly affected by system input characteristics and sampling interval size. Determination of a set of parameters

which might best be adjusted for improved performance may be aided by carrying out a sensitivity analysis of the system under consideration. This may also serve to reduce the number of parameters in the identification procedure and correspondingly reduce the computational task.

Alternate definitions of the weighting factors employed in the identification procedures can be made to permit tailoring of discrete state variables for some desired function. Through such a procedure distribution of the discretization error may be achieved to provide an improved response for some particular state or to gain improved performance over some interval of response time.

Further experimentation with discrete nonstationary systems may benefit from the segmental approximation approach. This approach requires less knowledge of system characteristics and broadens the applicability of the quasilinearization method.

Complete evaluation of the optimum discrete approximation technique and the parameter identification method can be made only through extended simulation studies of larger systems. Digital simulation of typical systems receiving attention in current research is unfortunately not practical with the present computer facility.

APPENDICES

APPENDIX I

DETERMINATION OF THE STEADY—STATE GAIN LOSS OF THE BLUM APPROXIMATION

The technique for development of a digital filter reported by Blum [25] was applied to the development of discrete approximations for continuous transfer functions by Fryer and Shultz [7]. In discussing the results obtained from a digital simulation employing this method, the authors note [7] that for the second—order system studied there was an unexpected reduction in the steady—state gain of the discretized system. A study of the discrete filter formulation presented by Blum reveals the basis for the reported gain loss.

Blum considers a linear time invariant filter with an input time sequence $x(nT)$ and an output time sequence $y(mT)$, which are related by the convolution summation

$$y(mT) = \sum_{n=0}^{\infty} w(\overline{m-n}\ T)\ x(nT), \tag{I.1}$$

where T is the sampling period, and w is the weighting sequence of the filter. This corresponds to Blum's Equation (18). However, in developing the expressions for the closed form of the digital filter, Blum redefines the weighting sequence by his Equation (27) to be

$$w(nT) \equiv Tw(nT). \tag{I.2}$$

Taking the z—transform of $y(mT)$, there results

$$Y(z) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} Tw(\overline{m-n}\ T)x(nT)z^{-m} \qquad (I.3)$$

Noting that $w(t)$ exists only for $t \geq 0$, and letting $k = m-n$, Equation (I.3) becomes

$$Y(z) = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} Tw(kT)x(nT)z^{-n-k}\ , \qquad (I.4)$$

which may be rewritten as

$$Y(z) = T \sum_{k=0}^{\infty} w(kT)z^{-k} \sum_{n=0}^{\infty} x(nT)z^{-n}\ . \qquad (I.5)$$

Recognizing the two summation expressions in Equation (I.5) as z-transforms of the filter weighting sequence $w(nT)$ and the filter input sequence $x(nT)$, and writing these as $G'(z)$ and $X(z)$, respectively, the expression for $Y(z)$ may be placed in the form

$$Y(z) = TG'(z)X(z) \qquad (I.6)$$

For the usual case of a sampled-data system with pulsed input $x(nT)$ and pulsed output $y(nT)$, the input and output are related by

$$Y(z) = G(z)X(z) \qquad (I.7)$$

where $G(z)$ is the pulse transfer function of the sampled-data system. Comparison of Equations (I.6) and (I.7) reveals that the discrete form of the system developed via Blum's technique will have a gain reduced by the magnitude of the sampling interval.

Equation (I.6) clearly shows that Blum's approach to

discretization of continuous filters is only a modification of the usual z—transfer function of a filter with pulsed input and output. For approximation of integrators of order n, the Blum technique produces the correct form of z—transfer function integrators since the z—transform of s^{-n} is multiplied by T to obtain the integrator expression. Approximation of continuous transfer functions other than integrators results however in the reduced gain factor observed in computational experiments by Fryer and Shultz and shown above.

APPENDIX II

Table 2

Examples of Optimum Discrete Operators

G(s)	n	$z^{-n}H_o(z)$	
		$R(s) = 1/s$	$R(s) = 1/s^2$
$\dfrac{1}{s}$	0	$\dfrac{Tz^{-1}}{1-z^{-1}}$	$\dfrac{T(1+z^{-1})}{2(1-z^{-1})}$
	1	$\dfrac{Tz^{-1}}{1-z^{-1}}$	$\dfrac{Tz^{-1}(4-3z^{-1}+z^{-2})}{2(1-z^{-1})}$
$\dfrac{1}{s^2}$	0	$\dfrac{T^2z^{-1}(1+z^{-1})}{2(1-z^{-1})^2}$	$\dfrac{T^2(1+4z^{-1}+z^{-2})}{6(1-z^{-1})^2}$
	1	$\dfrac{T^2z^{-1}(1+z^{-1})}{2(1-z^{-1})^2}$	$\dfrac{T^2z^{-1}(8-5z^{-1}+4z^{-2}-z^{-3})}{6(1-z^{-1})^2}$

Table 2 Continued

G(s)	n	$z^{-n}H_o(z)$	
		$R(s) = 1/s$	$R(s) = 1/s^2$
$\dfrac{a}{s+a}$	0	$\dfrac{(1 - e^{-aT})\,z^{-1}}{1 - e^{-aT}\,z^{-1}}$	$\dfrac{(aT - 1 + e^{-aT}) + z^{-1}(1 - e^{-aT} - aTe^{-aT})}{aT(1 - e^{-aT}\,z^{-1})}$
	1	$\dfrac{(1 - e^{-aT})\,z^{-1}}{1 - e^{-aT}\,z^{-1}}$	$\dfrac{z^{-1}d_1 + z^{-2}d_2 + z^{-3}d_3}{aT(1 - e^{-aT}\,z^{-1})}$

$$d_1 = 2aT - 1 + e^{-aT}$$

$$d_2 = 1 - aT + e^{-aT}(1 - 2aT) - 2e^{-2aT}$$

$$d_3 = e^{-aT}(aT - 1 + e^{-aT})$$

Table 2 Continued

G(s)		$z^{-n}H_o(z)$	
	n		
	0	$R(s) = 1/s$	$R(s) = 1/s^2$
1			$R(s) = 1/s$
	1	$\dfrac{z^{-1}(1 - e^{-aT}(\cos bT + a/b\sin bT)) + z^{-2}(e^{-2aT} - e^{-aT}(\cos bT - a/b\sin bT))}{(a^2 + b^2)(1 - z^{-1}2e^{-aT}\cos bT + z^{-2}e^{-2aT})}$	
$\dfrac{1}{(s+a)^2+b^2}$	0	$\dfrac{d_1 + z^{-1}d_2 + z^{-2}d_3}{T(a^2 + b^2)(1 - z^{-1}B + z^{-2}e^{-2aT})}$	$R(s) = 1/s^2$

$$d_1 = T(a^2 + b^2) + 2a(A-1)$$

$$d_2 = 2a(B + 1 - 2A - e^{-2aT}) - BT(a^2 + b^2)$$

$$d_3 = 2a(A - B) + e^{-2aT}(2a + T(a^2 + b^2))$$

Table 2 Continued

G(s)	n	$z^{-n} H_o(z)$	
		$R(s) = 1/s$	$R(s) = 1/s^2$

$$A = 2e^{-aT}\cos bT$$

$$B = e^{-aT}(\cos bT + (a^2 - b^2 / 2ab)\sin bT)$$

$$R(s) = 1/s^2$$

$$\frac{z^{-1}d_1 + z^{-2}d_2 + z^{-3}d_3 + z^{-4}d_4}{Tw_o^2(1 - z^{-1}A + z^{-2}e^{-2aT})}$$

$$w_o = a^2 + b^2$$

$$d_1 = 2Tw_o - 2a + AC - e^{-4aT}$$

$$d_2 = 2a(A + 1) - Tw_o(2A + 1) - 2aCe^{-2aT} - 4a(Ac - e^{-4aT})$$

1

Table 2 Continued

G(s)	n	$z^{-n}H_o(z)$	
		$R(s) = 1/s$	$R(s) = 1/s^2$

$d_3 = 2e^{-2aT}(Tw_o - a(1 - 2C)) + A(Tw_o - 2a) + 2a(AC - e^{-4aT})$

$A = 2e^{-aT}\cos bT$

$B = e^{-aT}(\cos bT - (a^2 - b^2 / 2ab) \sin bT)$

$C = e^{-aT}(\cos bT + (a^2 - b^2 / 2ab) \sin bT)$

APPENDIX III

EXAMPLE DEVELOPMENT OF AN OPTIMUM PULSE TRANSFER FUNCTION

Consider the second-order transfer function

$$G(s) = \frac{1}{(s + a)^2 + b^2}$$

for which the optimum discrete operator without delay is sought for a ramp input, i.e. $R(s) = 1/s^2$, and $F(z) = 1$. From these conditions are obtained the expressions

$$R(z) R(z^{-1}) F(z) F(z^{-1}) = \frac{T}{(1-z)^2} \frac{T}{(1-z^{-1})^2} \qquad (III.1)$$

and

$$A(z) = \frac{Tz}{(a^2 + b^2)(z-1)^2} - \frac{2az}{(a^2 + b^2)^2(z-1)}$$

$$+ \frac{2az^2 - 2ae^{-aT}z[\cos bT - (a^2 - b^2/2ab)\sin bT]}{(a^2 + b^2)[z^2 - z\ 2e^{-aT}\cos bT + e^{-2aT}]} \qquad (III.2)$$

- 110 -

From (III.1) are obtained

$$[R(z) \; R(z^{-1}) \; F(z) \; F(z^{-1})]_+ \; = \; \frac{T}{(1-z^{-1})^2}$$

and

$$[R(z) \; R(z^{-1}) \; F(z) \; F(z^{-1})]_- \; = \; \frac{T}{(1-z)^2} \; .$$

Evaluation of the numerator term of $H_o(z)$ requires obtaining the realizable part of the expression $zA(z)$. From Equation (III.2) the first term of $zA(z)$ is seen to be realizable while the realizable part of the second and third terms must be determined. Multiplying $A(z)$ by (z) and subtracting z from the second term gives for the realizable part of the resulting term

$$\frac{2a}{(a^2 + b^2)^2 \; (1-z^{-1})} \; .$$

Similarly treating the last term of the $A(z)$ expression yields for that term

$$\frac{2a}{(a^2+ b^2)^2} \; \frac{e^{-aT} [\cos bT + (a^2- b^2/2ab) \; \sin bT \,] \; - z^{-1}e^{-2aT}}{1 - z^{-1}2e^{-aT}\cos bT + z^{-2}e^{-2aT}} \; .$$

Combining the realizable terms and dividing by $T / (1-z^{-1})^2$ yields for the desired pulse transfer function

$$H_o(z) = \frac{d_1 + z^{-1}d_2 + z^{-2}d_3}{T(a^2 + b^2)^2 (1-z^{-1}B + z^{-2}e^{-2aT})} \quad ,$$

when

$$d_1 = T(a^2 + b^2) + 2a(A-1)$$

$$d_2 = 2a(B + 1 - 2A - e^{-2aT}) - BT(a^2 + b^2)$$

$$d_3 = 2a(A-B) + e^{-2aT}(2a + T(a^2 + b^2))$$

$$B = 2e^{-aT}\cos bT$$

$$A = e^{-aT}(\cos bT + (a^2 - b^2 /2ab) \sin bT) \quad .$$

The development of the final expression for $H_o(z)$ is facilitated by the partial fraction expansion of $A(z)$ in Equation (III.2).

APPENDIX IV

Flow Chart A

START

READ
INPUT

CALL
COMPUTE EXACT
SOLUTION

COMPUTE
DIFFERENCE
EQUATION
X(NT)

COMPUTE
NSES

WRITE
X
NSES

NO   READ
ALL   YES
DATA   STOP
?
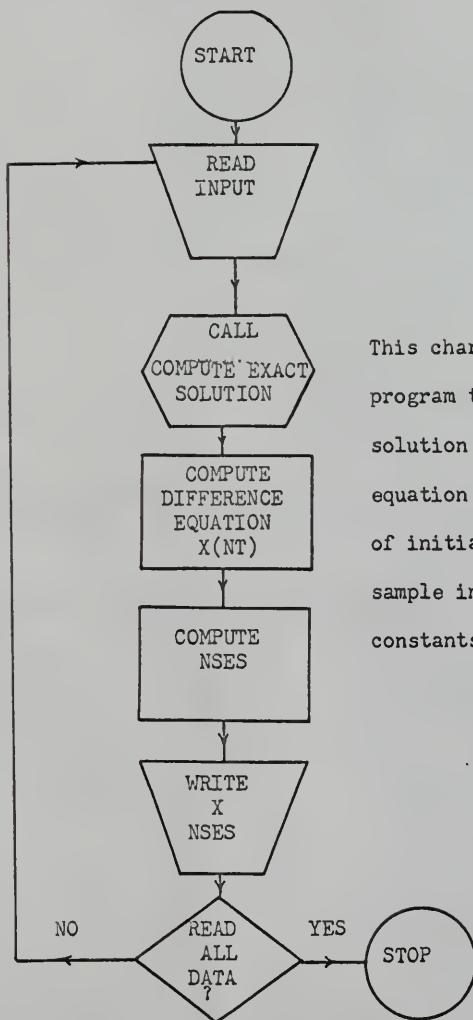
This chart describes a
program to compute the
solution to a difference
equation for a given set
of initial conditions,
sample interval and
constants.

- 113 -
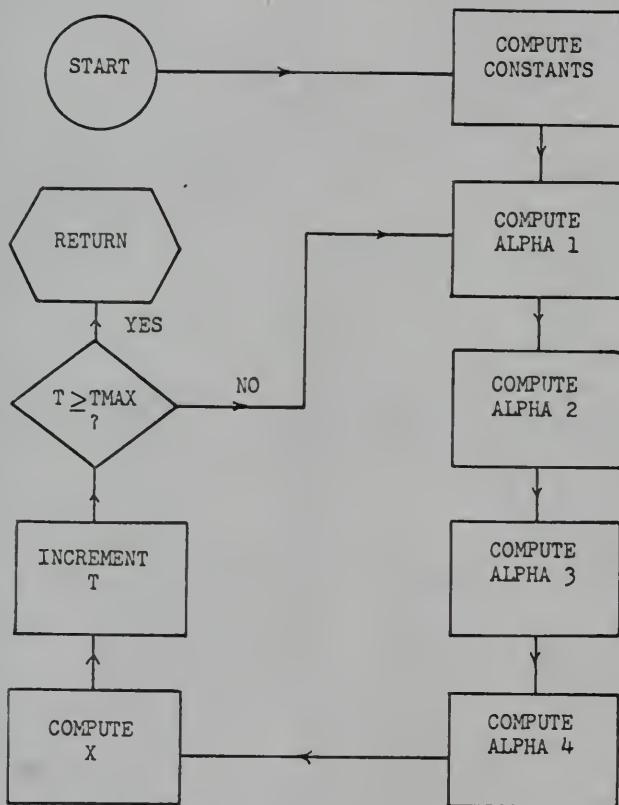
Flow Chart B

The subroutine described by this chart is an implementation of a fourth-order Runge-Kutta integration method. Initial conditions brought into the subroutine permit calculation of the solution for $\dot{x} = f(x,t)$.

Flow Chart C
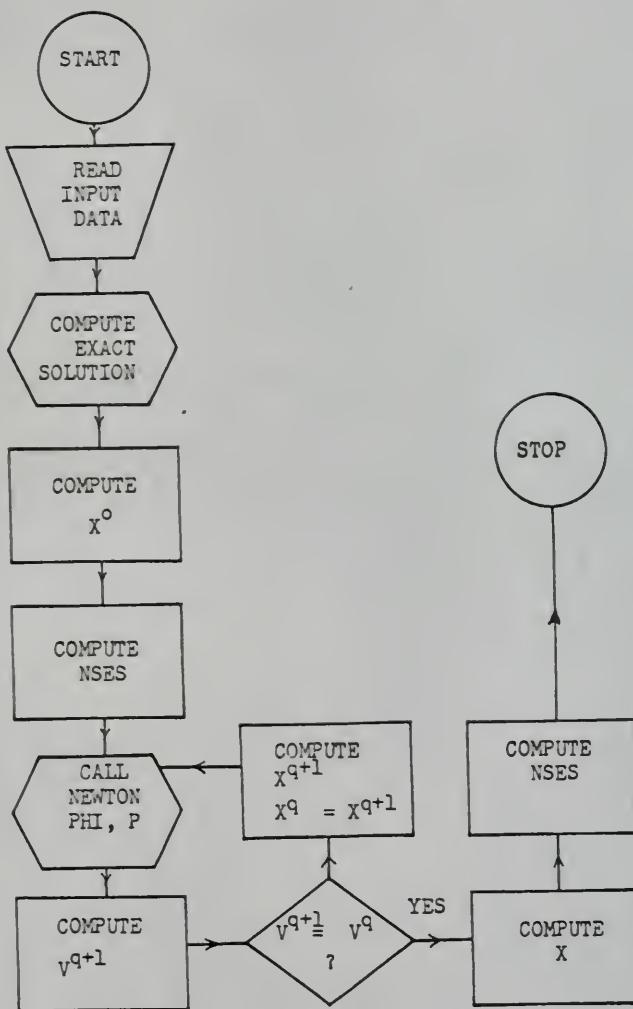
This chart diagrams the control sequence for a quasilinearization program.

Flow Chart D
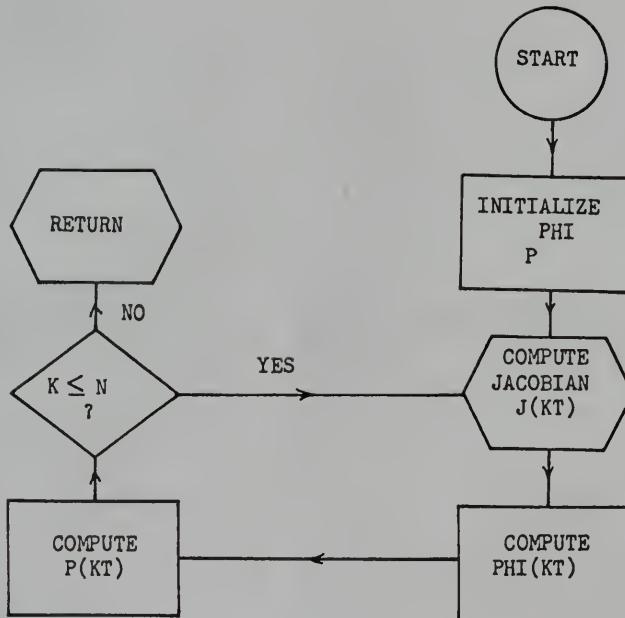
The NEWTON subroutine iterates the matrix manipulation to obtain the terminal homogeneous and particular solutions of the linearized equations for quasilinearization.

Flow Chart E
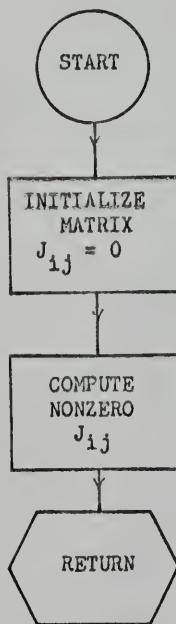
The subprogram diagrammed here computes the terms of the Jacobian matrix on each call.

Flow Chart F
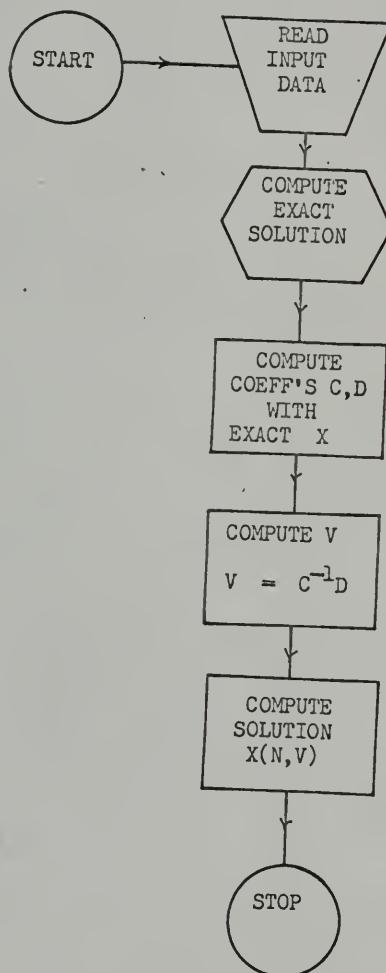
The program described by this chart implements the differential approximation method for parameter identification.

REFERENCES

1. Hamming, R. W., <u>Numerical Methods for Scientists and Engineers</u>, McGraw-Hill Book Company, Inc., New York, 1962.

2. Hildebrand, F. B., <u>Introduction to Numerical Analysis</u>, McGraw-Hill Book Company, Inc., New York, 1956.

3. Scarborough, J. B., <u>Numerical Mathematical Analysis</u>, The Johns Hopkins Press, Baltimore, 1950.

4. Tustin, A., "A method of analysing the behaviour of linear systems in terms of time series," <u>J.I.E.E.</u>, V 94, Pt. II-A May, 1947.

5. Fowler, M. C., "A new numerical method for simulation," <u>Simulation</u>, Vol 4, pp 324-330, May, 1965.

6. Boxer, R., "A note on numerical transform calculus," <u>Proc. IRE</u>, Vol 45, pp 1401-1406, October, 1957.

7. Fryer, W. D. and Shultz, W. C., "A survey of methods for digital simulation of control systems," Cornell Aeronautical Laboratory Report No. XA - 1681-F-1, July, 1964.

8. Hurt, J. M., "New difference equation technique for solving nonlinear differential equations," <u>AFIPS Conference Proceedings</u>, Vol 25, pp 169-179, 1964.

9. "Numerical techniques for real-time digital flight simulation," IBM Manual E20-0029-1, 1964.

10. Sage, A. P. and Burt, R. N., "Optimum design and error analysis
    of digital integrators for discrete system simulation,"
    AFIPS Conf. Proc., Vol 27, Part 1, pp 903-914, 1965.

11. Sage, A. P., "A technique for the real-time digital simulation
    of nonlinear control processes," Proc. of IEEE Region 3
    Conference, April 1966.

12. Bellman, R. and Kalaba, R., Quasilinearization and Nonlinear
    Boundary-Value Problems, American Elsevier Publishing
    Company, Inc., New York, 1965.

13. Bellman, R., Kagiwada, H., and Kalaba, R., "A computational
    procedure for optimal system design and utilization,"
    The Rand Corporation, RM-3174-PR, June 1962.

14. Warga, J., "On a class of iterative procedures for solving
    normal systems of ordinary differential equations,"
    Journal of Mathematics and Physics, Vol 29, pp 223-243 ,
    January 1953.

15. Bellman, R., "A note on differential approximation and orthogonal
    polynomials," The Rand Corporation, RM-3482-PR, March 1963.

16. Bellman, R., Kalaba, R., and Kotkin, B., "Differential
    approximation applied to the solution of convolutional
    equations," The Rand Corporation, RM-3601-NIH, May 1963.

17. Kalman, R. E. and Bertram, J. E., "A unified approach to the
    theory of sampling systems," J. Franklin Inst., Vol 267,
    pp 405-436, May 1959.

18. Zadeh, L. A. and Desoer, C. A., Linear System Theory, McGraw-Hill
    Book Company, Inc., New York, 1963.

19. Bekey, G. A., "Analysis and synthesis of discrete-time systems," <u>Modern Control Systems Technology</u>, Chapter 11, McGraw-Hill Book Company, Inc., New York, 1965.

20. Freeman, H., <u>Discrete-time Systems</u>, John Wiley and Sons, New York, 1965.

21. Hurewicz, W., "Filters and servo systems with pulsed data," <u>Theory of Servomechanisms</u>, M.I.T. Rad. Lab. Ser., Vol 25, Chapter 5, McGraw-Hill Book Company, Inc., New York, 1947.

22. Ragazzini, J. R. and Franklin, G. F., <u>Sampled-Data Control Systems</u>, McGraw-Hill Book Company, Inc., New York, 1958.

23. Tou, J. T. <u>Digital and Sampled-Data Control Systems</u>, McGraw-Hill Book Company, Inc., New York, 1959.

24. Jury, E. I., <u>Theory and Application of the z-Transform Method</u>, John Wiley and Sons, Inc., New York, 1963.

25. Blum, M., "Recursion formulas for growing memory digital filters," <u>IRE Trans. on Information Theory</u>, Vol 4, pp 24-30, March, 1958.

26. Tustin, <u>op. cit.</u>

27. Madwed, A., "Number series method of solving linear and nonlinear differential equations," M.I.T. Instrumentation Laboratory Report No. 6445-T-26, Cambridge, Mass., April, 1950.

28. Truxal, J. G. , "Numerical analysis for network design," <u>IRE Trans. on Circuit Theory</u>, Vol. CT-1, pp 49-60, September, 1954.

29. Boxer and Thaler, S., "A simplified method of solving linear and nonlinear systems," <u>Proc. IRE</u>, vol 44, pp 89-101, January, 1956.

30. Anderson, W. H., Ball, R.B. and Voss, J.R., "A numerical
    method for solving differential equations on digital
    computers," J. Assoc., Comp. Mach., Vol 7, pp 61-68,
    January,1960.

31. Fowler, M.E., "An example showing the use of root locus
    techniques to study nonlinear systems," IBM Systems Research
    and Development Center, Technical Report, August 12, 1964.

32. Burt, R.W., Optimum Design and Error Analysis of Digital
    Integrators, Unpublished item, University of Arizona,
    Tuscon, 1963.

33. Newton, G. C., Gould, L. A., and Kaiser, J. F., Analytical
    Design of Linear Feedback Controls, John Wiley and Sons,
    Inc., New York, 1957.

34. "An introduction to real-time digital flight simulation,"
    IBM Manual E20-0034-0, 1964.

35. Merriam, C. W., Optimization Theory and Design of Feedback Control
    Systems, McGraw-Hill Book Company, Inc., New York, 1964.

36. Breuer, M. A., "Digitalization of continuous control systems,"
    Simulation, V 5, pp 329-337, November, 1965.

37. Sage, A. P. and Melsa, J. L., "Optimum finite memory sampled-data
    systems," DP 63-718, I.E.E.E., May 1963.

38. Bellman, R., Gluss, B., and Roth, R., "Segmental differential
    approximation and the black box problem," The Rand Corporation,
    RM-4269-PR, October 1964.

39. Lavi, A. and Strauss, J. C., "Parameter identification in
    continuous control systems," IEEE International Convention

Record, Pt. 6, pp 49-61, March, 1963.

40. Detchmendy, D. M. and Sridhar, R., "On the experimental determination of the dynamical characteristics of physical systems," Proceedings of the National Electronics Conference, 1965, pp 575-580.

41. Bellman, R., "Successive approximations and computer storage problems in ordinary differential equations," Comm. of the ACM., Vol 4, pp 222-223, 1961.

42. McGill, R. and Kenneth, P., "Solution of variational problems by means of a generalized Newton-Raphson operator," AIAA Journal, Vol 2, pp 1761-1766, October, 1964.

43. Henrici, P., Discrete Variable Methods in Ordinary Differential Equations, John Wiley and Sons, Inc., New York, 1962.

44. Sylvester, R. J. and Meyer, F., "Two-point boundary-value problems by quasilinearization," J. Soc. Indust. Appl. Math, Vol 13, pp 586-602, June, 1965.

45. Sridhar, R., et al., "Investigation of optimization of attitude control systems," JPL Report TR-EE65-3, Purdue School of Electrical Engineering, Lafayette, Indiana, January, 1965.

46. Sage, A. P., "Optimal control theory," Lecture Notes, University of Florida, 1965.

47. Tou, J. T., Modern Control Theory, McGraw-Hill Book Company, Inc., New York, 1964.

Additional References

Adams, R. K., "Digital computer analysis of closed loop systems
     using the number series approach," Trans. AIEE, Vol 80,
     Pt II, pp 370-378, January, 1962.

Cuenod, M., "Methods de calcul a l'aide de suites," Imprimerie
     de la Concorde, Lausanne, 1955.

Salzer, J. M., "Frequency analysis of digital computers operating
     in real time," Proceedings IRE, vol 40, pp 457-466,
     February, 1954.

Mc Cormick, J. M. and Salvadori, M. G. Numerical Methods in
     Fortran, Prentice-Hall, Inc., Englewood Cliffs, New Jersey,
   1964.

BIOGRAPHICAL SKETCH

Stanley Louis Smith was born in Bagdad, Santa Rosa County, Florida, on March 22, 1934. He attended Santa Rosa County Public Schools and was graduated from Milton High School, Milton, Florida, in June, 1952. He entered the University of Florida in February, 1953, and received the degree of Bachelor of Electrical Engineering from this University in February, 1957.

After graduation he was employed by the Fort Worth Division of the General Dynamics Corporation for three years as a Test Engineer in the Engineering Test Laboratory. He entered the Graduate School of the University of Florida in September, 1960 and received the degree Master of Science in Engineering in January, 1962. During the course of this graduate program in electrical engineering he has been employed as a graduate teaching and research assistant in the Department of Electrical Engineering.

Mr. Smith is a member of the Institute of Electrical and Electronics Engineers and of Phi Eta Sigma, Sigma Tau and Phi Kappa Phi honorary societies.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Engineering and the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1966
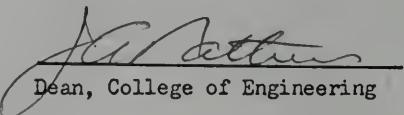
Dean, College of Engineering

Dean, Graduate School

Supervisory Committee:

Chairman

Co-Chairman